Schizophrenia, Human Leukocyte Antigen (HLA), and Herpes Viruses: Immunogenetic Associations at the Population Level

Neuroscience Insights
Volume 18: 1–22
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26331055231166411
SAGE

Lisa M James[1,2,3], Spyros A Charonis[1,2] and Apostolos P Georgopoulos[1,2,3,4]

[1]The HLA Research Group, Department of Veterans Affairs Health Care System, Brain Sciences Center, Minneapolis, MN, USA. [2]Department of Neuroscience, University of Minnesota Medical School, Minneapolis, MN, USA. [3]Department of Psychiatry, University of Minnesota Medical School, Minneapolis, MN, USA. [4]Department of Neurology, University of Minnesota Medical School, Minneapolis, MN, USA.

ABSTRACT: Several factors have been implicated in schizophrenia (SZ), including human herpes viruses (HHV) and the adaptive immunity Human Leukocyte Antigen (HLA) genes. Here we investigated these issues in 2 complementary ways. In one analysis, we evaluated SZ-HLA and HHV-HLA associations at the level of a single allele by computing (a) a SZ-HLA protection/susceptibility (P/S) score based on the covariance between SZ and 127 HLA allele prevalences in 14 European countries, (b) estimating in silico HHV-HLA best binding affinities for the 9 HHV strains, and (c) evaluating the dependence of P/S score on HHV-HLA binding affinities. These analyses yielded (a) a set of 127 SZ-HLA P/S scores, varying by >200× (maximum/minimum), which could not be accounted for by chance, (b) a set of 127 alleles × 9 HHV best-estimated affinities, varying by >600×, and (c) a set of correlations between SZ-HLA P/S scores and HHV-HLA binding which indicated a prominent role of HHV1. In a subsequent analysis, we extended these findings to the individual person by taking into account the fact that every individual carries 12 HLA alleles and computed (a) the average SZ-HLA P/S scores of 12 randomly chosen alleles (2 per gene), an indicator of HLA-based SZ P/S for an individual, and (b) the average of the corresponding HHV estimated affinities for those alleles, an indicator of overall effectiveness of HHV-HLA binding. We found (a) that HLA protection for SZ was significantly more prominent than susceptibility, and (b) that protective SZ-HLA scores were associated with higher HHV-HLA binding affinities, indicating that HLA binding and subsequent elimination of several HHV strains may confer protection against schizophrenia.

KEYWORDS: Human leukocyte antigen, HHV1, schizophrenia, immunogenetics, epidemiology

RECEIVED: October 24, 2022. **ACCEPTED:** March 13, 2023.

TYPE: Original Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Partial funding for this study was provided by the University of Minnesota (the Kunin Chair in Women's Healthy Brain Aging, the Brain and Genomics Fund, the McKnight Presidential Chair of Cognitive Neuroscience, and the American Legion Brain Sciences Chair) and the U.S. Department of Veterans Affairs. The sponsors had no role in the current study design, analysis or interpretation, or in the writing of this paper. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Lisa M James, Brain Sciences Center (11B), Minneapolis VAHCS, One Veterans Drive, Minneapolis, MN 55417, USA. Email: lmjames@umn.edu

Introduction

Schizophrenia is a complex and highly disabling mental disorder characterized by heterogeneous symptoms including hallucinations, delusions, disorganized speech, blunted affect, avolition, and anhedonia with a lifetime prevalence of ~1%.[1] Although the prevalence of schizophrenia is relatively low, its early onset and low remission together with high disability and psychosocial impairment result in a substantial global burden.[2] Accordingly, decades of research have been aimed at identifying risk and vulnerability factors to reduce the burden associated with this debilitating condition. Those efforts have documented a large genetic component to schizophrenia, with heritability estimated at around 80%.[3] In addition, other factors have also been implicated in schizophrenia, including various infections as well as other environmental factors (eg, obstetric complications, birth season, cannabis use, and adverse childhood experiences).[4-7] Indeed, considerable evidence supports a 2-hit theory that suggests that schizophrenia likely results from genetic vulnerability coupled with environmental hits during key neurodevelopmental periods.[6,8]

In light of extensive research implicating exposure to infections and genetic liability in schizophrenia risk, here we focus on the influence of genes involved in the human immune response to infections, human leukocyte antigen (HLA) genes, and human herpes viruses on schizophrenia risk.

HLA and schizophrenia

HLA genes, located on chromosome 6, code for cell-surface proteins that are instrumental in immune system surveillance and elimination of foreign (eg, viral, bacterial) antigens via 2 primary pathways that work in concert to provide host protection. Class I HLA molecules (HLA-A, B, C) are expressed on nucleated cells and bind small peptides from proteolytically degraded cytosolic viruses and bacteria; those bound peptides are exported to the cell surface for presentation to CD8+ cytotoxic T cells, signaling cell destruction. Class II HLA molecules (HLA-DPB1, DQB1, DRB1) are expressed on lymphocytes and professional antigen-presenting cells (eg, macrophages, dendritic cells, and monocytes). Class II HLA

presents larger peptides derived from endocytosed exogenous antigens to CD4+ T cells to facilitate B cell-mediated antibody production and adaptive immunity. More than 30 000 Class I and Class IIHLA alleles have been identified[9]; that variation maximizes population protection against infection by increasing the scope of antigens that can be bound and presented to the cell surface for destruction. Variation in HLA, particularly within the binding groove, has also been linked to variability in susceptibility to numerous human diseases.[10]

HLA is the most highly polymorphic region of the human genome,[11] differing not only across but also within populations.[12,13] The large number of polymorphisms coupled with the relatively low prevalence of schizophrenia and methodological variability across studies has resulted in inconsistent and modest associations in HLA candidate gene studies of schizophrenia[14]; however, early genome-wide association studies (GWAS) of schizophrenia have also documented associations with the HLA region,[15-17] and subsequent evidence has continued to support that association.[18-21] Furthermore, common HLA variants have been shown to influence brain abnormalities associated with schizophrenia including cerebral ventricle size,[22] thalamic volume and asymmetry,[23] and hippocampal volume,[24] and variations in the HLA binding groove have been shown to influence treatment response with antipsychotic medications in patients with schizophrenia.[25] Indeed, HLA-schizophrenia associations have been one of the most commonly reported genetic associations in the literature,[26] though effect sizes are relatively small and findings have varied across studies due to variation between samples and geographical variation in HLA frequency.[27]

An epidemiological approach to HLA-schizophrenia associations

HLA confers protection via pathogen elimination; consequently, HLA has evolved in concert with pathogen evolution[28,29] and has been shown to vary by population,[30,31] presumably reflecting differences in pathogen exposure. HLA-disease associations have been typically investigated using case-control studies.[32] Although this design is solid, the overall results have been rather limited for at least 2 reasons. First, such studies are frequently underpowered due to difficulties in obtaining large numbers of participants, with a notable exception of the study of HLA and malaria in West Africa.[33] And second, the drawback of a small sample size is compounded by the insistence on finding HLA alleles with a very low *P*-value. Given that diseases are common outcomes of various factors, including pathogen insults, host responses, molecular mimicry, autoimmunity, etc., in various combinations, the objective should rather be to identify and determine *the degree of involvement of different alleles in a given disease* rather than the quest to find "the one" allele with a very low *P*-value in disease association. Such an allele, when found, would be ranked #1 and would be most probably due to a specific aspect of the disease (eg, autoimmunity) but other alleles could play smaller but still important roles (protective or susceptibility) in disease manifestation, thus effectively modulating the effect of the most prominent allele. Hence, a method that combines signed, weighted contributions of individual alleles for an overall protection/susceptibility assessment would seem more appropriate. Finally, GWAS has been used in recent studies to investigate HLA-disease associations. A limitation of this approach, as noted elsewhere,[34] is that HLA alleles are been detected through imputation rather than direct HLA sequencing. While more time- and cost-effective than direct sequencing, the accuracy of HLA imputation for certain alleles (eg, DRB1) has been reported to be as low as 30%.[35] Conversely, due to the highly polymorphic nature of HLA, direct sequencing of large enough samples to identify HLA-disease associations is often prohibitive.

Here we take an alternative immunogenetic epidemiology approach that takes advantage of the population heterogeneity of HLA and permits the identification of a high-resolution population-level HLA profile with regard to disease prevalence. Using this approach, we[36-45] and others[30,46] have identified HLA alleles that are negatively associated with disease prevalence (ie, protective alleles) as well as HLA alleles that are positively associated with disease prevalence (ie, susceptibility alleles) in various neurological diseases,[36-41] type 1 diabetes,[42] and 30 types of cancer.[43-45] In light of HLA's involvement in pathogen elimination and autoimmunity, we presume that protective alleles facilitate pathogen elimination and that susceptibility alleles promote autoimmunity. Here we used this immunogenetic epidemiological approach above to evaluate associations between HLA frequency and schizophrenia prevalence and to derive a weighted overall measure along a protection-susceptibility continuum. Our working hypothesis is based on the following 3 assumptions: *(1) a number of HLA alleles are involved in SZ, (2) the contribution of an allele to SZ can be negative (protective) or positive (susceptibility), and (3) the ultimate contribution of the HLA to SZ in a particular individual lies in the combined contributions of 12 HLA alleles carried by the individual*, namely 2 of each 6 HLA classical genes (genes A, B, C of Class I and DPB1, DQB1, DRB1 of Class II); other nonclassical HLA genes (eg, DRB3, DRB4, etc.), not carried by all individuals, may contribute as well. The covariance of SZ prevalence and HLA allele frequency is a suitable quantitative estimate of the protective/susceptibility (P/S) contribution of an allele. Covariance possesses the following useful properties: (a) it can be easily calculated from the epidemiological/immunogenetic data (SZ prevalence and allele frequency), (b) it takes negative and positive values, corresponding to a protective or susceptibility contribution, (c) covariance estimates from different alleles can be added to derive and overall SZ P/S contribution of an *ensemble of alleles* as in the case of adding contributions of 12 alleles carried by an individual, (d) it can be normalized after division by the product of the data analyzed to yield a correlation coefficient, and, finally, (e) can be calculated from ranked data to provide a unit-free estimate of SZ

Table 1. Schizophrenia prevalence in the 14 CWE countries.

	COUNTRY	N SCHIZOPHRENIA	N TOTAL POPULATION	SCHIZOPHRENIA PREVALENCE (%)
1	Austria	30 182	8 800 000	0.3430
2	Belgium	37 116	11 300 000	0.3285
3	Denmark	16 343	5 700 000	0.2867
4	Finland	17 425	5 500 000	0.3168
5	France	205 206	64 600 000	0.3177
6	Germany	271 954	82 600 000	0.3292
7	Greece	35 645	10 800 000	0.3300
8	Italy	224 454	60 600 000	0.3704
9	Netherlands	76 112	17 000 000	0.4477
10	Norway	17 555	5 200 000	0.3376
11	Portugal	36 207	10 300 000	0.3515
12	Spain	155 797	43 300 000	0.3598
13	Sweden	33 169	9 900 000	0.3350
14	Switzerland	29 960	8 400 000	0.3567

P/S score. Here we use both the raw covariance and its normalized version (ie, correlation coefficient) to investigate (a) the SZ P/S contributions of various alleles and their 12-allele sets (the number of alleles carried by all individuals), and (b) the possible dependence of SZ prevalence on in silico estimated binding affinities of HLA alleles to 9 strains of human herpes virus.

Human herpes viruses (HHV) and schizophrenia

Infections, especially involving prenatal or childhood exposure, have been linked to adult schizophrenia risk purportedly via direct effects of pathogens and/or the effects of an inflammatory response to pathogens or autoimmunity on the developing brain.[47-49] A number of pathogens have been implicated in schizophrenia including several human herpes viruses (HHV).[50] HHV are neurotropic viruses that are nearly ubiquitous and can cause lifelong infection characterized by alternating periods of latency and reactivation.[51-53] They include herpes simplex virus 1 (HSV1/HHV1) and HHV2 (HHV2), varicella zoster virus (HHV3), Epstein-Barr virus (EBV/HHV4), cytomegalovirus (CMV/HHV5), HHV6a, HHV6b, HHV7, and HHV8. Though several studies have implicated HHV in schizophrenia, findings across studies have been highly inconsistent.[50,54] Nonetheless, intriguing links between HHV and schizophrenia exist. For instance, increased HHV antibodies have been documented in patients experiencing an initial episode of schizophrenia,[55,56] linked to higher suicide risk in patients with schizophrenia,[57] and have been associated with structural brain changes[56] and cognitive impairment[58,59] among patients with schizophrenia. In addition, studies have documented that anti-herpes treatment and anti-inflammatory medications improve schizophrenia symptoms.[55,60-62] Furthermore, studies suggest that interactions between single nucleotide polymorphisms in the HLA region and exposure to HHV influence schizophrenia risk.[19,63,64]

Consistent with studies documenting an interaction between the HLA region and HHV on schizophrenia,[19,63,64] we have hypothesized that exposure to foreign antigens, such as HHV, in the absence of HLA alleles that are capable of binding with those antigens can lead to antigen persistence, inflammation and/or autoimmunity, and downstream deleterious effects on the brain.[65-68] Here we extend that line of research to schizophrenia, evaluating the relation between the frequency of HLA alleles and the prevalence of schizophrenia in 14 Continental Western European countries and evaluating the correlation between the HLA-schizophrenia profile and HLA-HHV binding affinity.

Methods

Prevalence of schizophrenia

Table 1 shows the population prevalence of schizophrenia in 2016 for each of the following 14 countries in Continental Western Europe (CWE): Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Norway, Spain, Sweden, and Switzerland. Specifically, the total number of people with schizophrenia in each of the 14 CWE countries was identified from the Global Health Data Exchange,[69] a publicly available catalog of data from the Global Burden of

Disease study, the most comprehensive worldwide epidemiological study of more than 350 diseases. The number of people with schizophrenia in each country was divided by the total population of each country in 2016[70] and expressed as a percentage. We have previously shown that life expectancy for these countries is virtually identical[36]; therefore, life expectancy was not included in the current analyses.

HLA alleles

We obtained the population frequency of 127 HLA alleles from 14 Continental Western European Countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Norway, Spain, Sweden, and Switzerland), as described elsewhere.[39] Those 127 alleles comprised 69 Class I and 58 Class II alleles, all of which were used to analyze the schizophrenia-HLA immunogenetic scores (see below). The alleles and their mean frequencies of the individual alleles are given in Tables 2 and 3 for HLA Classes I and II, respectively.

SZ-HLA protection/susceptibility scores

The PSCorr scores are the Fisher *z*-transformed correlations between HLA schizophrenia prevalence and HLA allele frequency (both log-transformed).[36]

$$\text{PSCorr} = r' = \text{atanh}(r) \tag{1}$$

The PSCov scores are simply the covariance between the log-transformed SZ prevalence and HLA allele frequency. The sign of both scores indicates protection (negative) or susceptibility (positive) to schizophrenia (P/S scores), whereas their absolute value indicates the strength of their P/S effect.

HHV proteins

The amino acid sequences of surface glycoproteins from HHV 1 to 8 (1, 2, 3, 4, 5, 6A, 6B, 7, 8) were retrieved from the UniprotKB database[71] and are given in the Appendix. Table 4 gives the details of these proteins and associated information regarding the number of *n*-mers used in the analyses of Class I and Class II alleles.

A sliding window approach[67,68] (Figure 1) was used to partition the whole sequence of each glycoprotein into *n*-mers for HHV-HLA analyses. Since the exact *n*-mer length for binding to HLA Class I and Class II molecules is not known (only its range), we used 4 *n*-mer lengths (2 per HLA Class) to cover their approximate optimal range, namely 9- and 10-mer for Class I and 15-and 22-mer for Class II.[72] For each *n*-mer, a set of subsequences was generated (number of subsequences = length of glycoprotein—*n*). For all *n*-mers (n = 9 and 10 for Class I, and n = 15 and 22 for Class II), subsequences were collected and queried in the IEDB database (www.iedb.org) in order to identify potential epitope peptides that are recognized by and bind to HLA Class I and II surface receptor proteins. IEDB queries were performed for each of the sliding-window aggregated sequences against HLA alleles available (see below). Binding affinity predictions were obtained using the NetMHCIIpan method.[73] For each *n*-mer, a binding affinity score was predicted and reported as a percentile rank by comparing the peptide's score against the scores of 5 million random *n*-mers selected from the SwissProt database [71]; smaller percentile ranks indicate higher binding affinity ("good binders"). For each HHV strain and HLA allele, the lowest percentile rank (LPR) was used as the highest estimated binding affinity measure for quantitative analyses. The distribution of LPR was skewed to the left (Figure 2A) in an exponential fashion and was log-transformed (equation (2)) to normalize it (Figure 2B), as illustrated in Figures 2 and 3, respectively.

$$\ln(\text{LPR}) = \log_e(\text{LPR}) \tag{2}$$

Of a total of 127 alleles available (Tables 2 and 3), 113 were used for the analyses involving HHVs because 14 HLA Class I alleles could not be modeled by the NetMHCIIpan method (A*29:01, A*33:03, A*36:01, B*14:01, B*15:17, B*15:18, B*35:02, B*35:08, B*39:06, B*41:01, B*41:02, B*44:05, B*45:01, B*47:01).

Data analysis

Schizophrenia-HLA PSCorr scores. Standard parametric (mean, standard deviation, etc.) and nonparametric statistics (median, interquartile range, etc.) were used to evaluate the distribution of scores. We used Tukey's[74] fences to identify outlier and extreme values of the PSCorr distribution, as follows.

$$\text{Negative (protective) Inner Fence} = Q\_1 - 1.5\,\text{IQR} \tag{3}$$

$$\text{Negative (protective) Outer Fence} = Q\_1 - 3.0\,\text{IQR} \tag{4}$$

$$\text{Positive (susceptibility) Inner Fence} = Q\_3 + 1.5\,\text{IQR} \tag{5}$$

$$\text{Positive (susceptibility) Outer Fence} = Q\_3 + 3.0\,\text{IQR} \tag{6}$$

$$\text{Interquartile range: IQR} = Q\_3 - Q\_1 \tag{7}$$

where Q1, Q3 are the 25th and 75th percentiles, respectively. These fences demarcated outlier values (between inner and outer fences) and extreme values (outside the outer fences). A major objective of this study was to determine whether there was a statistically significant preponderance of protective or susceptibility schizophrenia PSCorr scores overall and within (and among) HLA classes and genes. For that purpose, we used a one-sample t-test with test value = 0 and estimated the mean preponderance, its statistical significance, and the effect size.[75,76]

Table 2. The 69 HLA Class I alleles used and their mean frequencies.

CLASS I								
GENE A			GENE B			GENE C		
	Allele	Frequency		Allele	Frequency		Allele	Frequency
1	A*01:01	0.1170	1	B*07:02	0.1009	1	C*01:02	0.0370
2	A*02:01	0.2715	2	B*08:01	0.0791	2	C*03:03	0.0506
3	A*02:05	0.0122	3	B*13:02	0.0238	3	C*04:01	0.1349
4	A*03:01	0.1501	4	B*14:01	0.0091	4	C*05:01	0.0716
5	A*11:01	0.0527	5	B*14:02	0.0275	5	C*06:02	0.0829
6	A*23:01	0.0237	6	B*15:01	0.0544	6	C*07:01	0.1472
7	A*24:02	0.1051	7	B*15:17	0.0104	7	C*07:02	0.1020
8	A*25:01	0.0139	8	B*15:18	0.0043	8	C*07:04	0.0146
9	A*26:01	0.0356	9	B*18:01	0.0609	9	C*12:02	0.0160
10	A*29:01	0.0058	10	B*27:02	0.0070	10	C*12:03	0.0678
11	A*29:02	0.0315	11	B*27:05	0.0435	11	C*14:02	0.0231
12	A*30:01	0.0165	12	B*35:01	0.0690	12	C*15:02	0.0370
13	A*30:02	0.0132	13	B*35:02	0.0187	13	C*16:01	0.0303
14	A*31:01	0.0295	14	B*35:03	0.0261			
15	A*32:01	0.0368	15	B*35:08	0.0111			
16	A*33:01	0.0116	16	B*37:01	0.0136			
17	A*33:03	0.0066	17	B*38:01	0.0276			
18	A*36:01	0.0063	18	B*39:01	0.0146			
19	A*68:01	0.0353	19	B*39:06	0.0069			
20	A*68:02	0.0220	20	B*40:01	0.0474			
			21	B*40:02	0.0212			
			22	B*41:01	0.0087			
			23	B*41:02	0.0056			
			24	B*44:02	0.0623			
			25	B*44:03	0.0431			
			26	B*44:05	0.0054			
			27	B*45:01	0.0090			
			28	B*47:01	0.0043			
			29	B*49:01	0.0220			
			30	B*50:01	0.0164			
			31	B*51:01	0.0640			
			32	B*52:01	0.0158			
			33	B*55:01	0.0129			
			34	B*56:01	0.0075			
			35	B*57:01	0.0278			
			36	B*58:01	0.0141			

Table 3. The 58 HLA Class II alleles used and their mean frequencies.

CLASS II							
GENE DPB1			GENE DQB1			GENE DRB1	
	Allele	Frequency		Allele	Frequency	Allele	Frequency
1	DPB1*01:01	0.0462	1	DQB1*02:01	0.1514	1 DRB1*01:01	0.0862
2	DPB1*02:01	0.1486	2	DQB1*02:02	0.0664	2 DRB1*01:02	0.0202
3	DPB1*02:02	0.0127	3	DQB1*03:01	0.1925	3 DRB1*01:03	0.0067
4	DPB1*03:01	0.1273	4	DQB1*03:02	0.1030	4 DRB1*03:01	0.1233
5	DPB1*04:01	0.3905	5	DQB1*03:03	0.0444	5 DRB1*04:01	0.0688
6	DPB1*04:02	0.1206	6	DQB1*04:02	0.0370	6 DRB1*04:02	0.0118
7	DPB1*05:01	0.0232	7	DQB1*05:01	0.1181	7 DRB1*04:03	0.0124
8	DPB1*06:01	0.0162	8	DQB1*05:02	0.0432	8 DRB1*04:04	0.0248
9	DPB1*09:01	0.0117	9	DQB1*05:03	0.0314	9 DRB1*04:05	0.0217
10	DPB1*10:01	0.0176	10	DQB1*06:01	0.0155	10 DRB1*04:07	0.0099
11	DPB1*11:01	0.0178	11	DQB1*06:02	0.1138	11 DRB1*04:08	0.0074
12	DPB1*13:01	0.0174	12	DQB1*06:03	0.0679	12 DRB1*07:01	0.1090
13	DPB1*14:01	0.0164	13	DQB1*06:04	0.0339	13 DRB1*08:01	0.0384
14	DPB1*17:01	0.0212	14	DQB1*06:09	0.0078	14 DRB1*08:03	0.0048
15	DPB1*19:01	0.0099				15 DRB1*09:01	0.0186
						16 DRB1*10:01	0.0124
						17 DRB1*11:01	0.0709
						18 DRB1*11:02	0.0076
						19 DRB1*11:03	0.0096
						20 DRB1*11:04	0.0498
						21 DRB1*12:01	0.0166
						22 DRB1*13:01	0.0713
						23 DRB1*13:02	0.0433
						24 DRB1*13:03	0.0130
						25 DRB1*13:05	0.0034
						26 DRB1*14:01	0.0226
						27 DRB1*15:01	0.1104
						28 DRB1*15:02	0.0115
						29 DRB1*16:01	0.0342

Random permutations test for assessing the statistical significance of the schizophrenia-HLA profile. The SZ-HLA profile is a vector of length N = 127 containing the estimated PSCorr scores. We tested the null hypothesis that the assignment of PSCorr to specific alleles may be due to chance by performing the permutation test below. Let H be the original (observed) SZ-HLA profile and H' be the profile obtained by random permutation, as follows. (1) For each allele, its observed frequencies were paired to randomly permuted prevalences of the 14 CWE countries. (2) The Pearson correlation between allele frequency and corresponding (to the permuted country) SZ prevalence was computed. (3) These correlations were Fisher z-transformed to yield a "permuted" SZ-HLA profile, H'. (4) Next, we computed the sum of the absolute paired differences

Table 4. Attributes of the 9 HHV viruses used.

VIRUS	PROTEIN	UNIPROTKB ID	PROTEIN LENGTH (AA)
HHV-1	Envelope glycoprotein D	Q69091	394
HHV-2	Envelope glycoprotein D	P03172	393
HHV-3	Envelope glycoprotein E	Q9J3M8	623
iHHV-4	Envelope glycoprotein B	P03188	857
HHV-5	Envelope glycoprotein B	P06473	906
HHV-6A	Glycoprotein Q2	P0DOE0	214
HHV-6B	Glycoprotein Q1	Q9QJ11	516
HHV-7	Envelope glycoprotein H	P52353	690
HHV-8	Envelope glycoprotein H	F5HAK9	730

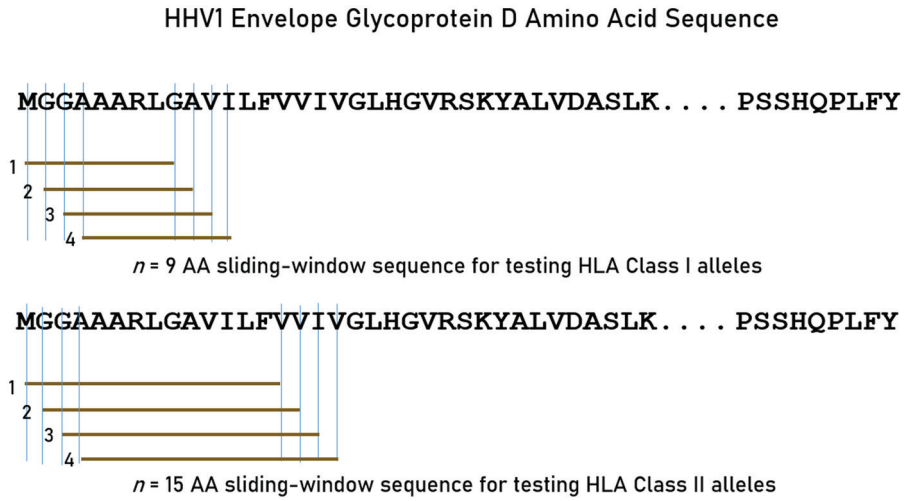

Figure 1. Schematic diagram to illustrate the sliding window approach for estimating exhaustively in silico the binding affinity of all consecutive 9- and 15-mer peptides. The figure refers to HHV1 for illustration but the glycoproteins of all HHV strains (Table 4) were tested, as described in the text.

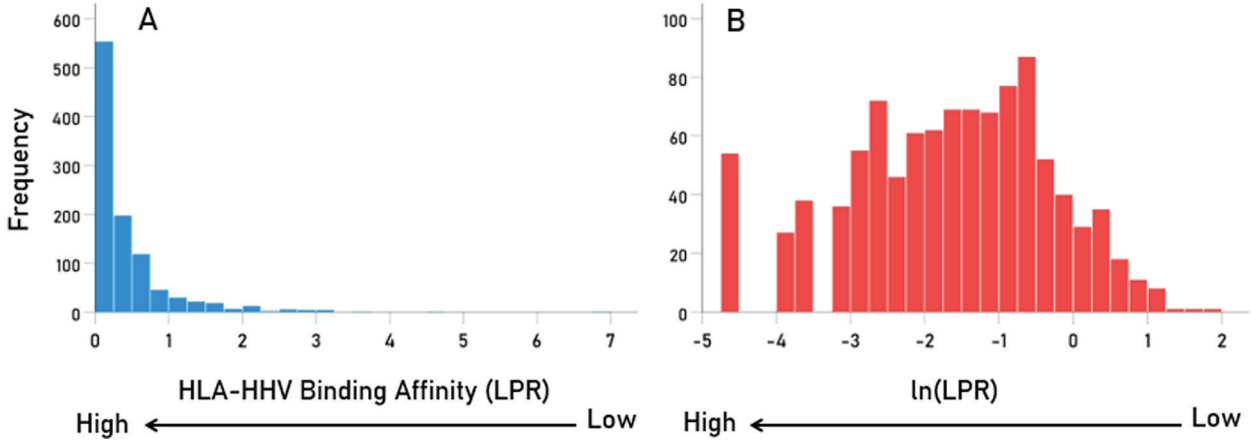

Figure 2. Frequency distributions of LPR [A] and ln(LPR) [B]. N = 113 alleles × 9 HHV strains = 1017 for each distribution.

between H and H': a sum of zero would indicate that the 2 profiles are the same. (5) Finally, we carried out this procedure 1 000 000 times and counted the number of times M that that sum was equal to zero, indicating that the randomly obtained profile would be the same as the observed one. Then, the ratio $w = \frac{M}{1,000,000}$ is the probability that the observed profile H could be due to chance. In a variant of this test, we determined in 1000 runs how frequently the rank of the PSCorr scores

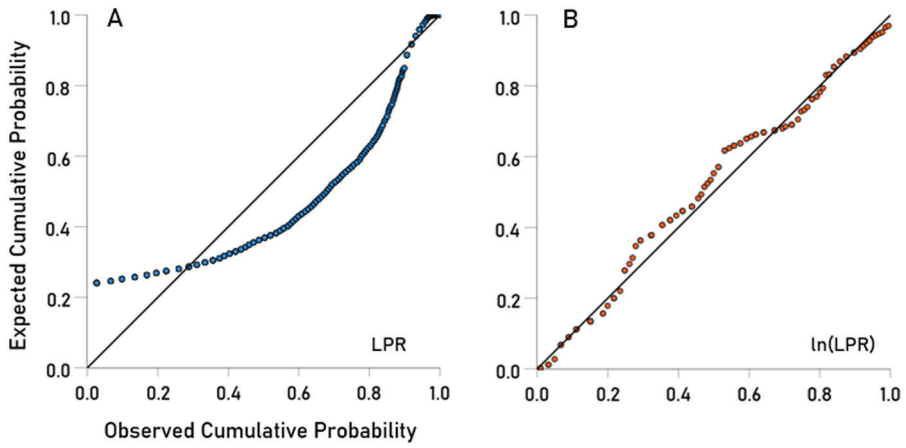

Figure 3. Probability-probability plots of LPR [A] and $\ln(\text{LPR})$ [B] to illustrate the substantial departure of the LPR distribution from normal (diagonal) and conversion to close to normal when log-transformed.

could occur randomly, thus relaxing the requirement of an exact match.

Application to individuals. Since every individual carries 12 classical HLA alleles (2 of each 3 HLA Class I genes and 2 of each 3 Class II genes), an average PSCorr score can be calculated:

$$\begin{aligned}\xi &= \overline{\text{PSCorr}} = [r'(A_1) + r'(A_2) + r'(B_1) + r'(B_2) \\ &+ r'(C_1) + r'(C_2) + + r'(DPB1_1) + r'(DPB1_2) \\ &+ r'(DQB1_1) + r'(DQB1_2) + r'(DRB1_1) \\ &+ r'(DRB1_2)]/12\end{aligned} \tag{8}$$

Finally, we transformed $\overline{\text{PSCorr}}$ (an average of weighted r') to the original Pearson correlation scale, so that its range would be [−1, +1]:

$$\xi = \tanh(\overline{\text{PSCorr}}) \tag{9}$$

We obtained expected estimates of ξ using a bootstrap procedure,[77] as follows. For each gene, 2 r' scores were drawn randomly (with replacement) from the pool of available alleles of that gene and averaged to yield a bootstrap ξ*, for a simulated "individual." The procedure was repeated 1 million times for a total of 1 000 000 ξ*. In the resulting distribution, we identified negative (protective) and positive (susceptibility) outlier and extreme values of ξ*, as described above (equations (3)–(7)). The percentage of positive (susceptibility) outliers was compared to the percentages of schizophrenia prevalence above (Table 1) using an independent samples t-test.

Application to populations. In this analysis, we extended computations above to populations by using the average allele frequencies (Table 2) as weighting factors in equation (8):

$$\begin{aligned}\Xi &= \overline{\text{PSCorr}'} = [r'(A_1) f(A_1) + r'(A_2) f(A_1) \\ &+ r'(B_1) f(B_1) + r'(B_2) f(B_2) + r'(C_1) f(C_1) \\ &+ r'(C_2) f(C_2) + + r'(DPB1_1) f(DPB1_1) \\ &+ r'(DPB1_2) f(DPB1_2) + r'(DQB1_1) f(DQB1_1) \\ &+ r'(DQB1_2) f(DQB1_2) + r'(DRB1_1) f(DRB1_1) \\ &+ r'(DRB1_2) f(DRB1_2)]/12\end{aligned} \tag{10}$$

We also carried out a bootstrap analysis of Ξ (N = 1 000 000) to obtain 1 million bootstrap Ξ* values for evaluating the representation in the population of allele frequency-weighted combinations of randomly selected 12 SZ PSCorr scores.

Association of schizophrenia PSCorr scores with HHV affinities. The second objective of this study was to assess the relation between schizophrenia PSCorr scores and the in silico estimated binding affinity of the corresponding alleles with the 9 HHVs (Table 4). For this purpose, we performed 2 analyses. At the single allele level, we carried out a stepwise linear regression, where the schizophrenia PSCorr score was the dependent variable and the HLA-HHV binding affinities, ln(LPR) (equation (2)), of the 9 HHV strains were the independent variables. At the level of the individual, we took into account the fact that an individual carries 12 HLA alleles, each of which has an estimated binding affinity to the 9 HHV strains. We first computed the average of those affinities for each allele and then computed the grand average binding affinity for the whole set of the randomly chosen alleles in the bootstrap procedure above. This provided a set of 1 million pairs consisting of (a) the average PSCorr (ξ*) of the 12 randomly chosen alleles, and (b) the grand average (ζ*) of the estimated HHV affinities for that set of alleles. The correlation between ξ* and ζ* is an estimate of the aggregate association of SZ PSCorr and HHV at the level of the individual.

Schizophrenia-HLA PSCov scores. The same analyses as above were performed on PSCov scores.

Implementation of analysis procedures. The IBM-SPSS statistical package (version 27) was used for implementing standard statistical analyses, including nonparametric exploratory data analysis,[74] regression analysis, and testing of proportions. All *P* values reported are 2-sided. The permutation test and bootstrap procedure were implemented using FORTRAN (Geany, version 1.38, built on or after 2021-10-09) and 64-bit Mersenne Twister random number generator with a large random double-precision odd seed.

Results

The SZ-HLA immunogenetic scores PSCorr and PSCov are given in Tables 5 and 6. The 2 scores were highly significantly and positively correlated (Figure 4; $r = .937$, $P = 3.94 \times 10^{-59}$, N = 127).

Permutation tests

In the permutation test, no cases were found where a random SZ prevalence – HLA allele frequency pairing (out of 1000 000 runs) matched the observed SZ-HLA PSCorr or PSCov profiles (Tables 5 and 6), thus rejecting the null hypothesis that the observed profile could be accounted for by chance ($P < 1 \times 10^{-6}$) for both PSCorr and PSCov scores. In addition, in the ranks version of the random permutation test, which relaxed the requirement of an exact PSCorr match, we found that in only 3/1000 runs the ranks of the observed PSCorr scores matched the ranked observed SZ-HLA PSCorr profile, thus rejecting again the null hypothesis that the observed ranks of the PSCorr scores were due to chance ($P = .003$); no case matching the PSCov scores profile was found, the rejecting the null hypothesis that the observed ranks of the PSCov score were due to chance ($P < .001$).

Frequency distributions

The ranked SZ-HLA PSCorr scores are plotted in Figure 5, and their frequency distribution and box plot are shown in Figure 6A and B, respectively; detailed statistics are given in Table 7. It can be seen that there were 3 outliers (in red), with scores lower that the lower inner fence (equation (3), −0.081). These were HLA alleles B*27:02, B*35:08, and DRB1*13:05, illustrated in Figure 7, with SZ-HLA PSCorr scores of −0.956, −0.870, and −0.820, respectively. The correlation coefficients were all >.5 ($r = .743$ for B*27:02, 0.703 for B*35:08, and 0.675 for DRB1:13:05), indicating large effect sizes.[76]

Very similar results were obtained for PSCov, with the following 2 differences. (a) There was an additional protective outlier allele (B*15:18), and (b) there were 4 outlier susceptibility alleles (C*16:01, DRB1*11:04, DRB1*11:01, DQB1*02:02—from highest to lower score). These results indicate a higher sensitivity of the PSCov measure.

Tests of score proportions

Since the sign of the PSCorr score is determined by the sign of PSCov score, the following results apply to both scores. (a) There was a statistically significant higher proportion of protective scores (77/127 = 60.6%; $P = .017$, Wilson test of single proportion). (b) The proportions of protective scores in either Class I or Class II, and in any of the 6 genes did not differ from those of the susceptibility scores.

Tests of score magnitude

With respect to PSCorr scores, a statistically significant higher magnitude of protective (vs susceptibility) scores was observed (a) overall, (b) for HLA Class I, and (c) for HLA B gene (Figure 8); no significant effects were found for Class II or any other gene. Detailed statistics for the significant effects are given in Tables 8 and 9. Similarly, the magnitude of PSCov scores was significantly higher overall, HLA Class I, and HLA B gene.

Application to individuals

The following results were obtained from the bootstrap analysis of the PSCorr scores (descriptive statistics in Table 10); very similar results were obtained for the PSCov scores (not illustrated). Since the SZ-HLA PSCorr scores are population-level correlations between schizophrenia prevalence and HLA allele frequencies, it is reasonable to interpret them as reflecting the degree of protection/susceptibility conferred in general on an individual of the population by their specific HLA genetic makeup. For a particular individual, the average of the PSCorr scores of the 12 HLA alleles that the individual carries (equation (8)) can serve as an overall estimate, ξ, of SZ protection/susceptibility along a continuum ranging from −1 to +1 (equation (9)). Analysis of the distribution of the bootstrapped 1 million ξ^* values revealed the following. (a) The distribution was unimodal, approximating a normal distribution but skewed to the left, that is, toward negative (protection) values (Figure 9). (b) Indeed, negative (protection) values (71.2%) outnumbered significantly positive (susceptibility) values (28.8%) ($P < .001$, Wilson score test for a single proportion). In addition, the mean of the distribution was negative, indicating a bias toward SZ protection ($P < .001$, one-sample *t*-test against the null hypothesis that the mean = 0). (c) There were 4404 (0.4%) negative (protective) ξ^* outliers and 3 extreme ones (>Tukey's outer fence, equation (4); Figures 10 and 11A). (d) There were 3043 (0.304%) positive (susceptibility outliers (Figures 10 and 11B). These proportions differed highly significantly ($Z = 15.8$, $P < .00001$). (e) Remarkably, the percent (0.3%) of positive (susceptibility) values of ξ^* above was very close

Table 5. Schizophrenia Immunogenetic scores for HLA Class I alleles.

CLASS I								
GENE A			GENE B			GENE C		
Allele	PSCorr	PScov	Allele	PSCorr	PSCov	Allele	PSCorr	PSCov
A*01:01	−.248	−.002528	B*07:02	.135	.006762	C*01:02	−.702	−.019900
A*02:01	−.183	−.002039	B*08:01	.078	.002951	C*03:03	−.119	−.006297
A*02:05	−.089	−.005197	B*13:02	.015	.000647	C*04:01	−.317	−.005400
A*03:01	−.200	−.007437	B*14:01	.078	.004697	C*05:01	.131	.005611
A*11:01	.081	.002187	B*14:02	−.011	−.000632	C*06:02	.467	.013424
A*23:01	.030	.001572	B*15:01	.168	.011421	C*07:01	−.142	−.001682
A*24:02	−.018	−.000413	B*15:17	−.304	−.017833	C*07:02	.309	.013948
A*25:01	.006	.000252	B*15:18	−.757	−.044797	C*07:04	−.066	−.003011
A*26:01	.110	.004525	B*18:01	−.195	−.011494	C*12:02	−.254	−.018436
A*29:01	−.490	−.028436	B*27:02	−.956	−.037332	C*12:03	−.107	−.005263
A*29:02	.158	.015396	B*27:05	−.301	−.018373	C*14:02	−.012	−.000765
A*30:01	.208	.012826	B*35:01	−.084	−.002405	C*15:02	.084	.003699
A*30:02	.136	.010422	B*35:02	−.450	−.026727	C*16:01	.399	.044524
A*31:01	.077	.003210	B*35:03	−.205	−.011519			
A*32:01	.017	.000581	B*35:08	−.872	−.051535			
A*33:01	−.154	−.010631	B*37:01	.344	.014528			
A*33:03	−.374	−.021057	B*38:01	−.323	−.009779			
A*36:01	−.058	−.003207	B*39:01	.074	.002330			
A*68:01	.148	.003772	B*39:06	−.102	−.004251			
A*68:02	−.094	−.008357	B*40:01	.004	.000311			
			B*40:02	−.437	−.017620			
			B*41:01	−.147	−.007591			
			B*41:02	−.503	−.025780			
			B*44:02	.075	.001952			
			B*44:03	.184	.011248			
			B*44:05	−.163	−.008413			
			B*45:01	−.026	−.001748			
			B*47:01	−.290	−.014689			
			B*49:01	−.059	−.004296			
			B*50:01	−.289	−.013632			
			B*51:01	.128	.019624			
			B*52:01	−.510	−.023853			
			B*55:01	.425	.014714			
			B*56:01	−.575	−.021447			
			B*57:01	.291	.010743			
			B*58:01	−.421	−.018861			

Table 6. Schizophrenia immunogenetic scores for HLA Class II alleles.

CLASS II								
GENE DPB1			GENE DQB1			GENE DRB1		
Allele	PSCorr	PSCov	Allele	PSCorr	PSCov	Allele	PSCorr	PSCov
DPB1*01:01	−.006	−.000296	DQB1*02:01	−.322	−.009983	DRB1*01:01	−.246	−.006472
DPB1*02:01	.146	.002918	DQB1*02:02	.368	.033053	DRB1*01:02	−.062	−.003773
DPB1*02:02	−.270	−.024760	DQB1*03:01	.249	.012999	DRB1*01:03	−.103	−.005160
DPB1*03:01	.194	.008266	DQB1*03:02	−.161	−.004341	DRB1*03:01	.403	.013356
DPB1*04:01	−.209	−.003374	DQB1*03:03	−.160	−.008421	DRB1*04:01	−0243	−.020587
DPB1*04:02	−.378	−.008933	DQB1*04:02	−.190	−.011956	DRB1*04:02	−.020	−.001756
DPB1*05:01	−.152	−.006469	DQB1*05:01	−.094	−.001957	DRB1*04:03	.028	.001474
DPB1*06:01	−.115	−.006798	DQB1*05:02	−.195	−.015455	DRB1*04:04	.252	.015393
DPB1*09:01	.305	.010258	DQB1*05:03	.468	.024950	DRB1*04:05	−.055	−.005655
DPB1*10:01	−.484	−.025886	DQB1*06:01	−.313	−.020175	DRB1*04:07	.200	0.016376
DPB1*11:01	.217	.018771	DQB1*06:02	−.217	−.011312	DRB1*04:08	−.365	−.030033
DPB1*13:01	−.336	−.015974	DQB1*06:03	−.112	−.003141	DRB1*07:01	.167	.008761
DPB1*14:01	−.153	−.006537	DQB1*06:04	.188	.007809	DRB1*08:01	−.172	−.012190
DPB1*17:01	.326	.018901	DQB1*06:09	.170	.004685	DRB1*08:03	−.373	−.016610
DPB1*19:01	.058	.004375				DRB1*09:01	−.012	−.001582
						DRB1*10:01	.146	.008958
						DRB1*11:01	.546	.034239
						DRB1*11:02	−.253	−.016144
						DRB1*11:03	−.009	−.000537
						DRB1*11:04	.369	.036578
						DRB1*12:01	−.138	−.005480
						DRB1*13:01	−.145	−.003224
						DRB1*13:02	.143	.003480
						DRB1*13:03	.106	.003680
						DRB1*13:05	−.820	−.039492
						DRB1*14:01	.008	.000502
						DRB1*15:01	−.192	−.008125
						DRB1*15:02	−.330	−.015932
						DRB1*16:01	−.093	−.007554

to, and statistically not significantly different from, the prevalences of schizophrenia used in this study (Table 1; $P = .318$, independent samples t-test).

Application to populations

In this analysis, we computed the average Ξ of SZ-HLA PSCorr scores weighted by the corresponding allele frequencies (equation 10) to obtain the population outcome of a particular combination of HLA alleles. The results obtained are shown in Table 11; very similar results were obtained for the SZ PSCov score (data not illustrated). More specifically: (a) The distribution of Ξ*was unimodal, approximating a normal distribution, with a skew to the left (ie, toward protection values). (b) Overall, negative (protection) values (69.3%) outnumbered significantly positive (susceptibility) values (30.7%) ($P < .001$, Wilson score

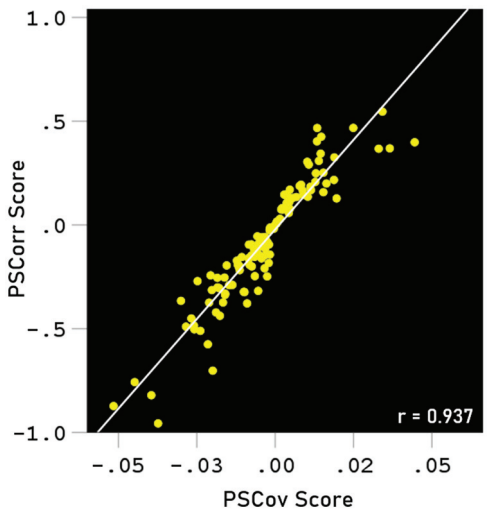

Figure 4. SZ-HLA PSCorr scores are plotted against SZ-PSCov scores. See text for details.

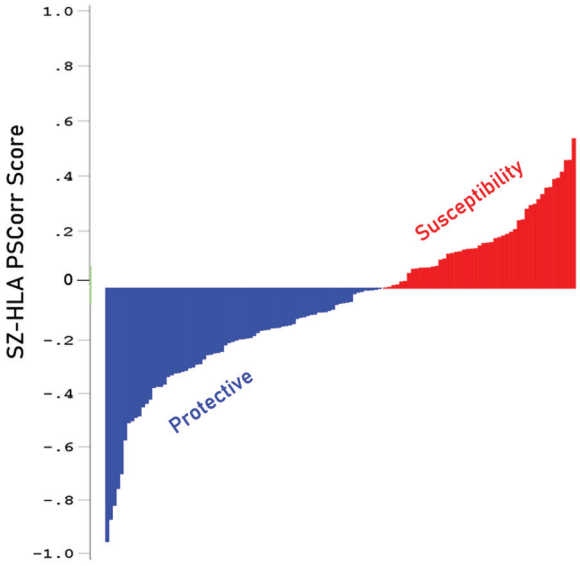

Figure 5. Ranked SZ-HLA PSCorr are plotted against their ranks. Notice that protective scores (blue) are more numerous and stronger than susceptibility scores (red). See text for details.

test for a single proportion). In addition, the mean of the distribution was negative, indicating a bias toward SZ protection ($P < .001$, one-sample t-test against the null hypothesis that the mean = 0). (c) There were 5562 (0.556%) negative (protective) Ξ^* outliers and 11 extreme ones (0.0011%). (d) There were 3134 (0.313%) positive (susceptibility outliers. These proportions differed highly significantly ($Z = 26.09$, $P < .00001$). (e) Similarly to the results of ξ^* above, the percent of susceptibility outliers was very close to, and statistically not significantly different from, the prevalences of schizophrenia used in this study (Table 1; $P = .434$, independent samples t-test).

Schizophrenia PSCorr scores and HLA-HHV binding

In this analysis, we first estimated the highest binding affinity of each HLA allele to the 9 HHV strains (Table 4). We found that HLA alleles of Class I had significantly better binding affinities than those of Class II (Figure 12, $P < .001$, F-test in ANOVA). The average binding affinities did not differ significantly among genes of the same HLA Class (Figure 13; $P = .884$ for Class I genes, and $P = .492$ for Class II genes, ANOVA). Next, we evaluated the possible dependence of SZ-HLA P/S scores (PSCorr and PSCov) on HHV-specific HLA-binding affinities by performing a stepwise multiple linear regression, where the SZ-HLA immunogenetic score of individual alleles was the dependent variable and the 9 corresponding HHV-specific ln(LPR) values (one per allele and HHV strain) were the independent variables. Preliminary analyses revealed that there were significant differences between the 2 HLA classes, so we performed the regression analysis above separately for HLA Class I (N = 55) and Class II (N = 58) alleles. Moreover, for each HLA Class, we performed the analysis twice, namely 1 per n-mer (9- and 10-mer for Class I, and 15- and 22-nmer for Class II). The results were clear-cut, as follows. (a) No statistically significant associations were found in any Class II analyses (15- or 22-mer) between SZ PSCorr scores and HHV ln(LPR) for any HHV strain. In

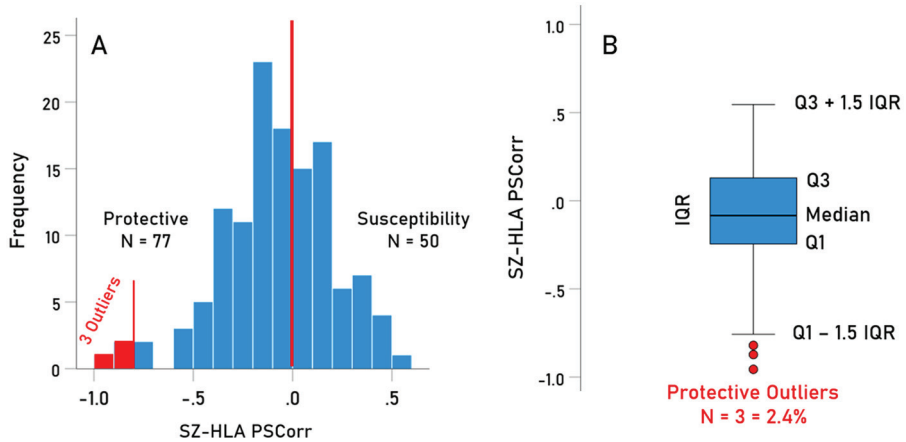

Figure 6. (A) frequency distribution of the 127 SZ-HLA PSCorr scores. (B) boxplot of the data in Figure 5. Q1, 25th percentile; Q3, 75th percentile; IQR, interquartile range.

Table 7. Descriptive statistics of the distribution of SZ-HLA PSCorr scores. N = 127 HLA alleles.

Mean	−0.0763
SEM	0.0249
Median	−0.0838
SD	0.2808
Minimum	−0.9559
Maximum	0.5463
Range	1.5023
IQR	0.3772
25th percentile (Q1)	−0.2463
75th percentile (Q3)	0.1308

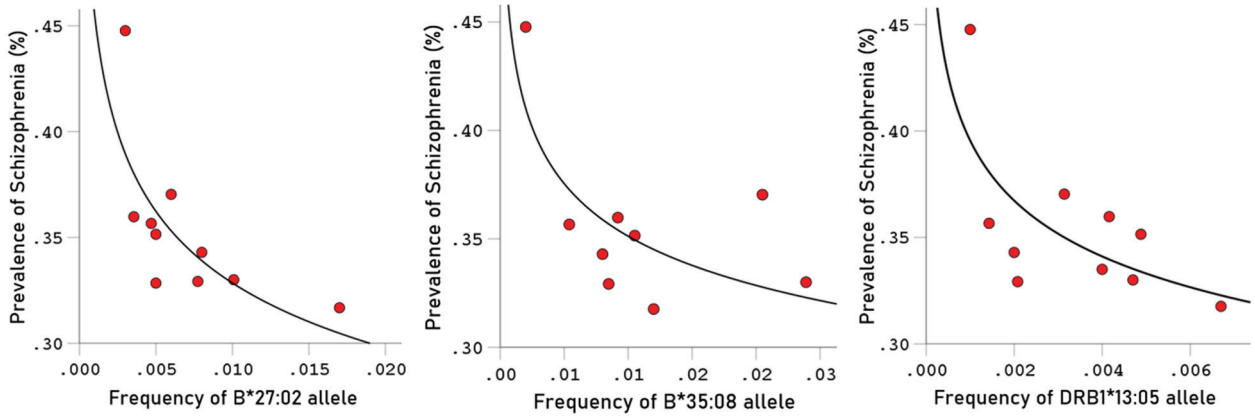

Figure 7. Prevalence of schizophrenia is plotted against frequency of B*27:02 ($r = .743$, PSCorr = −.956), B*35:01 ($r = -.703$, PSCorr = −.872), and DRB1*13:05 ($r = -.675$, PSCorr = −.820), as indicated. The curves are a power fits. The fits are linear between log-log transformed data, from which the correlations above are derived.

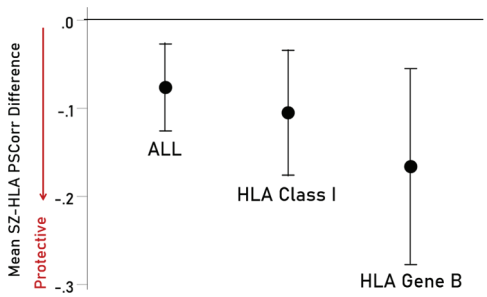

Figure 8. Mean (±95% CI) protective preponderance of SZ-HLA scores (negative/protective – positive/susceptibility SZ-HLA score) for the groups indicated. See text for details.

contrast, (b) a statistically significant association was found in both Class I analyses (9- and 10-mer) only between PSCorr scores and HHV1 ln(LRP) for both 9-mer (Figure 14A: $r = .333$, $P = .013$, N = 55) and 10-mer AA sequences (Figure 14B: $r = .339$, $P = .011$, N = 55); the effect sizes of these correlations were medium-to-large.[75,76] Practically the same associations were found between PSCov and HHV1 ln(LPR): (a) No statistically significant associations were found in any Class II analyses (15- or 22-mer) between SZ PSCov scores and HHV ln(LPR) for any HHV strain. In contrast, (b) a statistically significant association was found in both Class I analyses (9- and 10-mer) only between PSCov scores and HHV1 ln(LRP) for both 9-mer ($r = .330$, $P = .014$, N = 55) and 10-mer AA sequences ($r = .305$, $P = .023$, N = 55).

Combined SZ-HLA PSCorr scores versus combined HLA-HHV affinities

In the previous section, we analyzed the affinities of single HLA alleles to the various HHV strains. Since an individual carries 12 HLA alleles, it is reasonable to suppose that the ultimate influence of HHV on SZ susceptibility of an individual would be reflected in the correlation between the average PSCorr ξ (equation (8)) and the average estimated 9 HHV binding affinities ζ (equation (11)). For that purpose, we ran a bootstrap procedure (see Methods) and obtained 1 000 000 bootstrap values of ξ^* and ζ^*. We found that these measures were positively and highly significantly correlated ($r = .644$, $Z = 764.98$, $P < .001$). The data from a portion of that set

Table 8. Statistics for statistically significant preponderance of protective (negative) Schizophrenia-HLA PSCorr scores (one sample *t*-test, test value = 0).

	N	MEAN DIFFERENCE	95% CONFIDENCE INTERVAL		ITI	TWO-TAILED *P* VALUE
			LOWER	UPPER		
All	127	−0.0763	−0.1256	−0.0269	3.060	.003
Class I	69	−0.1050	−0.1758	−0.0342	2.958	.004
Gene B	36	−0.1661	−0.2773	−0.0549	3.034	.005

Table 9. Effect sizes for the results in Table 7.

	COHEN'S *D*	EFFECT SIZE	95% CONFIDENCE INTERVAL	
			LOWER	UPPER
Overall	−0.272	Small-to-Medium	−0.448	−0.094
Class I	−0.356	Small-to-Medium	−0.598	−0.112
Gene B	−0.506	Medium-to-Large	−0.850	−0.155

Table 10. Descriptive statistics of the distribution of ξ*. N = 1 000 000.

Mean	−0.00197
SEM	0.000005
Median	−0.00204
SD	0.0046
Minimum	−0.0246
Maximum	0.0208
Range	0.0455
IQR	0.006169
25th percentile (Q1)	−0.005092
75th percentile (Q3)	0.001077

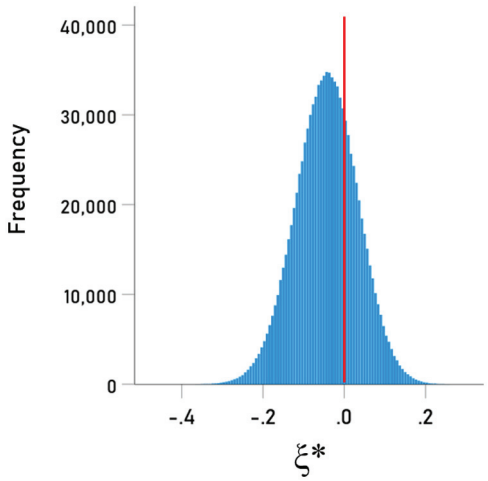

Figure 9. A, frequency distribution of 1 million bootstrap ξ* values.

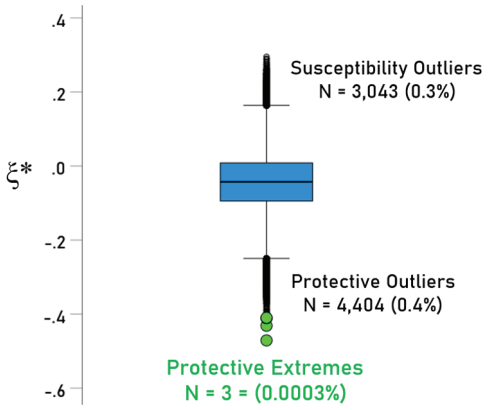

Figure 10. Boxplot of the data in Figure 9. Conventions as in Figure 6B.

(N = 1000) are plotted in Figure 15. Very similar results were obtained for PSCov (data not illustrated). These results indicate a strong dependence of overall individual SZ risk on combined HHV affinities, such that a HLA makeup with overall higher binding HHV affinity would confer protection from (ie, lower the risk of) schizophrenia.

Discussion

Methodological considerations

The SZ-HLA PSCorr and PSCov scores were used here as continuously-varying measures of schizophrenia-HLA allele covariation at the population level and not as genetic causative factors testing a specific null hypothesis; hence no statistical tests of significance or *P*-values were computed for individual PSCorr scores since there was no specific null hypothesis to be tested. Instead, we tested the null hypothesis that the *set* (vector) of the 127 SZ-HLA PSCorr scores was due to chance by performing an extensive random permutation test. The results of this test rejected the null hypothesis above ($P < 1 \times 10^{-6}$) as

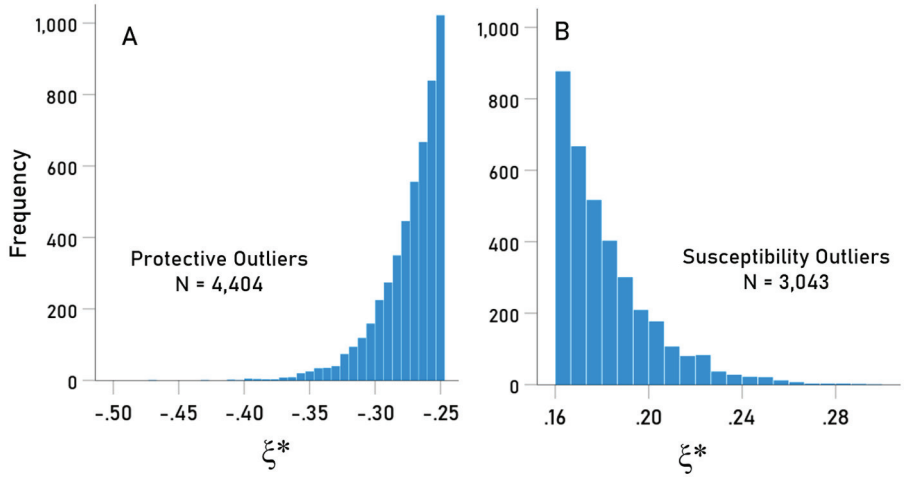

Figure 11. Frequency distribution of protective (A) and susceptibility (B) ξ* values.

Table 11. Descriptive statistics of the distribution of Ξ*. N = 1 000 000.

Mean	−0.03608
SEM	0.000071
Median	−0.03510
SD	0.0709
Minimum	−0.454
Maximum	0.295
Range	0.749
IQR	0.0946
25th percentile (Q1)	−0.08281
75th percentile (Q3)	0.01176

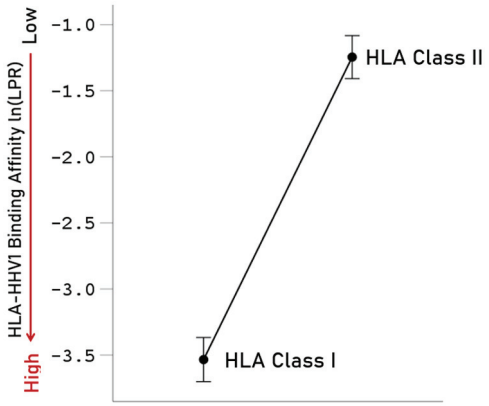

Figure 12. Average (±95% CI) HHV1-HLA affinities [ln(LPR)] for HLA Class I and II.

well as its more liberal version that the ranked scores were due to chance ($P < .003$). In accordance with these results, the set of the observed SZ-HLA PSCorr scores was employed to evaluate its possible associations with the in silico estimated binding affinities to HHV proteins, which were used in the same way, namely as continuously-varying quantitative assessments HLA-HHV associations.

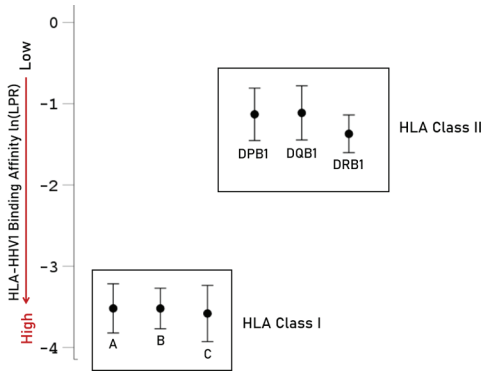

Figure 13. Average (±95% CI) HHV1-HLA affinities [ln(LPR)] for genes of HLA Class I and II.

Remarkably, but unsurprisingly, practically identical results were obtained in all analyses when SZ PSCov scores were used, instead of PSCorr scores, since these measures were strongly and highly significantly correlated (Figure 4). In a way, the PSCov score would be somewhat more appropriate because it measures pure covariation, without standardization, which, although useful for statistical significant testing, it alters the covariance estimate itself. It should be mentioned that covariance has been usefully[78,79] (and profitably[80]) employed in various fields, including evolution and natural selection[78,79] and finance.[80]

General

A wealth of research has documented genetic and environmental influences on schizophrenia, including exposure to infections. Here we employed an epidemiological-immunogenetic approach[30,36-46] to explore the relations between schizophrenia and HLA at the population level by computing the covariance (and its standardized version of correlation) between schizophrenia prevalence and HLA allele frequencies in 14 CWE

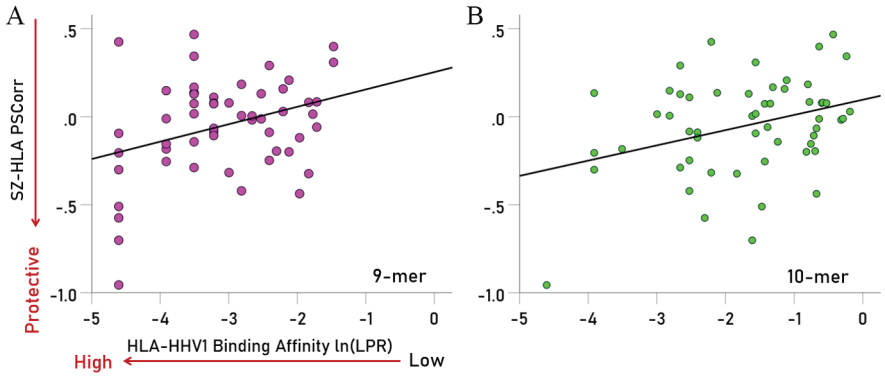

Figure 14. SZ-HLA PSCorr scores are plotted against ln(LPR) for HHV1 for 9-mer (A; left panel) and 10-mer (B; right panel) analyses. See text for details.

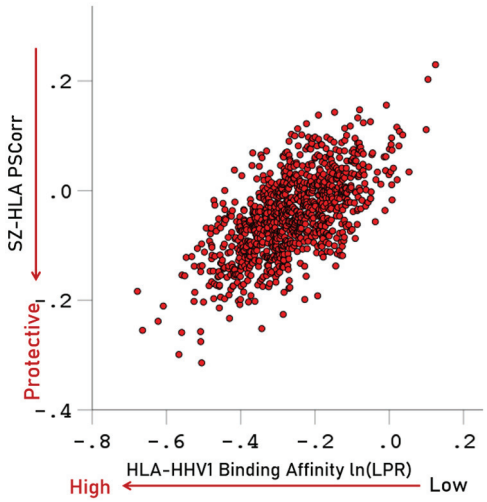

Figure 15. One-thousand ξ* values are plotted against the corresponding averages of 9 HHV ln(LPR) values. $r = .638$, $P < .001$. See text for details.

countries. The choice of these countries was based mainly on the availability of HLA allele frequencies and on their practically identical life expectancies. The computed SZ-HLA PSCorr (and PSCov) scores were used to (a) estimate the degree of protection (negative score) or susceptibility (positive score) conferred by specific HLA alleles on the individuals of the population, (b) evaluate in silico possible associations between those scores and the binding affinities of the same HLA alleles to proteins of 9 HHV strains, and (c) investigate the possible dependence of the SZ P/S scores on HHV-HLA binding affinities.. Our findings documented a preponderance of protective effects of HLA on schizophrenia prevalence and indicated that the SZ-HLA profile was highly associated with the estimated HLA binding affinity for HHV1, such that higher HLA-HHV1 binding affinity was associated with protective SZ-HLA scores. We then extended those analyses to aggregates of SZ-HLA scores and corresponding HHV-HLA binding affinities, since every individual carries 12 HLA alleles. Indeed, we found that at this aggregate level, SZ protection conferred by HLA was highly associated with better (more effective) binding affinities to HHV. Taken together, these findings suggest that HLA binding and subsequent elimination of HHV1 confer protection against schizophrenia in individuals and populations.

HLA and schizophrenia prevalence

With regard to the association of HLA and schizophrenia, we found there were more negative associations between HLA allele frequency and schizophrenia prevalence than there were positive associations, reflecting an overall protective effect. The relative protection conferred by HLA may partially account for the low prevalence of schizophrenia. Significant protective effects were documented here for Class I HLA, and the B genes, in particular. Notably, the HLA-B gene has been previously linked to the intergenerational transmission of schizophrenia risk.[81] Here, 63.9% of the HLA-B genes were protective against schizophrenia and, of the 127 alleles investigated, the most highly significant protective effect was found for HLA-B*27:02 ($r' = -0.956$), an allele that has been strongly linked to risk for ankylosing spondylitis in one of the strongest HLA-disease associations known.[82] Interestingly, this negative SZ-HLA association corresponds to the similarly negative associations between ankylosing spondylitis and schizophrenia documented in epidemiological studies.[83] Although there was a preponderance of protective (ie, negative) HLA-SZ associations, 39% of HLA-SZ associations were positive, suggesting susceptibility toward SZ. HLA is most commonly associated with immune-mediated/autoimmune conditions.[33] In light of substantial evidence indicating robust associations between the risk of schizophrenia and several autoimmune disorders[84-89]the present findings suggest HLA may mediate that association. Furthermore, it has been suggested that some schizophrenia endophenotypes are driven by an infectious/inflammatory and/or autoimmune component [27,62,90-93]; the present findings suggest that HLA may influence immune-mediated schizophrenia endophenotypes.

Implications for individuals

A major outcome of this study was the derivation of a unique, continuously-varying measure, ξ, assessing the protection/susceptibility of an individual to schizophrenia, based on their specific genetic HLA makeup. The allele-based SZ-HLA PSCorr scores provided the building blocks of ξ, which is a standardized score of the average 12 PSCorr scores of the HLA alleles carried by an individual (equations (8) and (9)). A reasonable interpretation of ξ would be as a measure of schizotypy,[94-97] such that high values of ξ would correspond with increased liability for schizophrenia, as supported by the finding that the percent of very high (outlier) positive (susceptibility) values of ξ* in our simulation was 0.3%, a percent very close to, and statistically not significantly different from, the prevalences of schizophrenia used in this study (Table 1). The same considerations apply to the negative (protective) ξ values which could be interpreted as schizophrenia-resilience[96] scores. Since ξ is a normalized score with a range [−1, +1], it can be used as a threshold for detecting individuals resilient or susceptible to schizophrenia. Specifically, a value of $\xi < -0.25$ (the value of negative inner fence, equation (3)) would indicate schizophrenia resilience, whereas a value of $\xi > 0.16$ (the value of positive inner fence, equation (5)) would indicate schizophrenia susceptibility, with continuous gradation in both domains.

SZ-HLA PSCorr scores and in silico HHV binding affinity

Several prior studies have documented the influence of HHV1 on schizophrenia. For instance, HHV1 antibodies are elevated in patients with schizophrenia compared to healthy controls[56,98] (cf, ref[99]), and HHV1 exposure is associated with brain morphological anomalies[56,61,100] and cognitive impairments[58,61,64,100-102] in patients with schizophrenia. Here, we did not evaluate the association between HHV and schizophrenia but rather the influence of HLA on both by examining the dependence of SZ-HLA immunogenetic scores on HHV-specific HLA-binding affinities. Those analyses revealed a significant association between the HLA-schizophrenia profile and binding affinity only for HHV1, bolstering evidence regarding the influence of HHV1 on schizophrenia. As reviewed elsewhere,[50] associations between other HHV including CMV, HHV2, and EBV with schizophrenia have been mixed. Here, at the allele level, SZ P/S scores were found to be associated with HHV1-HLA affinity. However, at the allele aggregate level, where groups of 12 alleles were considered (as carried by every individual), the association between SZ-HLA P/S scores and HHV-HLA binding affinities was much stronger, indicating the involvement of several HHV strains acting in concert. In fact, this finding is in accord with the central hypothesis of this study, namely that SZ risk/protection is the outcome of a combined effect of groups of HLA alleles and not the exclusive effect of a specific allele.

Considering the above findings as a whole, the present study suggests that the association between HHV and schizophrenia may be partially attributable to HLA composition. The formation of an antigen-HLA molecule complex is a vital step in antigen elimination and host protection against foreign antigens. Given the extremely polymorphic nature of HLA, there is large variability in the binding affinity to a given antigen across HLA alleles. In the case of poor binding affinity between certain HLA molecules and antigen epitopes, the elimination of those antigens is blocked resulting in antigen persistence which may result in inflammation, autoimmunity, and disease.[65] Indeed, we have previously discussed antigen persistence due to HLA incongruence in relation to several conditions.[65-68,103-106] We documented here that the HLA-schizophrenia profile that is primarily characterized by protective effects is significantly associated with HHV binding, suggesting that HLA binding and elimination of HHV protect against schizophrenia. The converse implication, which follows from the perspective of the persistent antigen theory, is that in the absence of HLA alleles capable of binding HHV antigens with sufficient affinity and immunogenicity, such antigens may persist,[107] ultimately contributing to damaging effects including those associated with schizophrenia. That is, the effect of HHV1 exposure may be partially mitigated by HLA with high binding affinity. Furthermore, the combined influence of HLA composition and HHV1 exposure may partly account for the low prevalence of schizophrenia despite high HHV1 seroprevalence in Europe.[108]

The current findings and approach may also have implications that extend beyond schizophrenia given the high degree of genetic overlap that exists between risk for schizophrenia and other psychiatric disorders.[109-111] For instance, previous studies have documented substantial genetic overlap between schizophrenia and bipolar disorder.[109-111] Furthermore, like schizophrenia, both infections (eg, HHV[112]) and HLA[113] have been implicated in bipolar disorder risk, and, like schizophrenia, anti-inflammatory agents have shown promising treatment effects for bipolar disorder.[114] The extent to which the immunogenetic profile for bipolar disorder and other psychiatric conditions overlaps with that of schizophrenia remains to be investigated and is an interesting avenue for future studies.

The "crowdfunding" nature of HLA contribution to schizophrenia (and other diseases)

In evaluating HLA contributions to schizophrenia, it is important to consider the following: (1) the HLA region is the most polymorphic of the entire human genome, (2) every individual possesses 12 HLA alleles, and (3) each HLA allele may differ in terms of the direction of association with schizophrenia (and other diseases) and the magnitude of those associations.

Consequently, rather than focusing on the individual "unique" contribution of specific alleles on schizophrenia susceptibility and protection, the primary emphasis here is on relatively small contributions of individual alleles that when combined as a set contribute to overall protection/susceptibility of schizophrenia in a population and for individuals. This "crowdfunding" nature of HLA contributions to schizophrenia moves away from single gene-disease associations toward considering polygenetic contributions to disease.

Conclusions

Overall the findings of the present study imply that schizophrenia prevalence is partially attributable to the combined effects of HLA composition and HHV1 exposure in Continental Western Europe. A significant strength of this study relative to GWAS studies that rely on imputation is that the approach used here permits evaluation of the frequency of high-resolution Class I and Class II HLA alleles with schizophrenia prevalence and estimated HHV binding affinity in silico. High-resolution HLA genotyping allows for the evaluation of HLA-disease associations at the protein-level where small variations may result in discordant disease associations such as that seen here for HLA-C*07:01 which was negatively correlated with schizophrenia prevalence and C*07:02 which was positively correlated with schizophrenia prevalence. In addition, the immunogenetic epidemiological approach used here takes advantage of HLA heterogeneity to evaluate the association of a large number of HLA alleles with schizophrenia at the population level, providing a broad picture of the influence of HLA on the population prevalence of schizophrenia. Given that schizophrenia is polygenic, our approach of combining signed, weighted contributions of HLA alleles to derive a unique, continuously-varying, normalized measure could easily be extended to include contributions of others genes for a more comprehensive estimate. At the minimum, determination of the complete, high-resolution HLA makeup of individuals in a population would be needed, in addition to population frequencies of other schizophrenia-related genes, and also mutations identified in GWAS studies.

Limitations of the study

Despite these strengths, it is also prudent to consider the present findings within the context of the study's limitations. First of all, these findings hold for the European countries studied, and their extension to other countries remains to be determined, given the geographic heterogeneity of schizophrenia prevalence,[2] HLA allele frequency,[30,31] and pathogen exposure.[115] Also, our study was limited to 127 HLA alleles, the only alleles that occurred in 9 or more of the countries investigated at the time the data was obtained; still, other alleles that were not included here may be relevant to schizophrenia and HHV, especially in countries of other geographical locations. An additional consideration is that this study was limited to HHV although several bacterial and other viral infectious agents have been implicated in schizophrenia.[4-7] Future studies evaluating the HLA-schizophrenia profile in relation to the binding affinity of HLA to other infectious agents may reveal additional HLA-mediated infectious contributors to schizophrenia.

Author Contributions

LMJ and SAC contributed to data retrieval. APG contributed to data analysis. LMJ and APG wrote the paper. LMJ, SAC, and APG edited and approved the paper.

Data Availability

The datasets generated and analyzed in the current study are publicly available.

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors.

ORCID iDs

Lisa M James https://orcid.org/0000-0001-9021-0690
Apostolos P Georgopoulos https://orcid.org/0000-0003-4412-725X

Significance Statement

Schizophrenia prevalence covaries with the population frequencies of human leukocyte antigen (HLA) alleles, which, in turn, bind with various affinities to human herpes viruses. HLA alleles protective against schizophrenia had higher binding affinities to HHV, suggesting that HLA binding and subsequent elimination of HHV may confer protection against schizophrenia.

REFERENCES

1. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med*. 2005;2:e141.
2. Charlson FJ, Ferrari AJ, Santomauro DF, et al. Global epidemiology and burden of schizophrenia: findings from the Global Burden of Disease Study 2016. *Schizophr Bull*. 2018;44:1195-1203.
3. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187-1192.
4. Brown AS. The environment and susceptibility to schizophrenia. *Prog Neurobiol*. 2011;93:23-58.
5. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: update 2012. *Mol Psychiatry*. 2012;17:1228-1238.
6. Davis J, Eyre H, Jacka FN, et al. A review of vulnerability and risks for schizophrenia: beyond the two hit hypothesis. *Neurosci Biobehav Rev*. 2016;65:185-194.
7. Robinson N, Bergen SE. Environmental risk factors for schizophrenia and bipolar disorder and their relationship to genetic risk: current knowledge and future directions. *Front Genet*. 2021;12:686666.
8. Bayer TA, Falkai P, Maier W. Genetic and non-genetic vulnerability factors in schizophrenia: the basis of the "two hit hypothesis". *J Psychiatr Res*. 1999;33:543-548.
9. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res*. 2020;48:D948-D955.
10. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol*. 2018;18:325-339.
11. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet*. 2013;14:301-323.

12. Heath SC, Gut IG, Brennan P, et al. Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet*. 2008;16:1413-1429.
13. Buhler S, Nunes JM, Nicoloso G, Tiercy JM, Sanchez-Mazas A. The heterogeneous HLA genetic makeup of the Swiss population. *PLoS One*. 2012;7:e41400.
14. Wright P, Nimgaonkar VL, Donaldson PT, Murray RM. Schizophrenia and HLA: a review. *Schizophr Res*. 2001;47:1-12.
15. Purcell SM, Wray NR, Stone JL. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748-752. doi:10.1038/nature0818
16. Shi J, Levinson DF, Duan J, et al. Common variants on chromosome 6p22. 1 are associated with schizophrenia. *Nature*. 2009;460:753-757. doi:10.1038/nature 08192
17. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744-747. doi:10.1038/nature08186
18. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969-976. doi:10.1038/ng.940
19. Bamne M, Wood J, Chowdari K, et al. Evaluation of HLA polymorphisms in relation to schizophrenia risk and infectious exposure. *Schizophr Bull*. 2012;38: 1149-1154.
20. Irish Schizophrenia Genomics Consortium and the Wellcome Trust Case Control Consortium 2. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol Psychiatry*. 2012;72:620-628.
21. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;520:177-183. doi:10.1038/ nature16549
22. Agartz I, Brown AA, Rimol LM, et al. Common sequence variants in the major histocompatibility complex region associate with cerebral ventricular size in schizophrenia. *Biol Psychiatry*. 2011;70:696-698.
23. Brucato N, Guadalupe T, Franke B, Fisher SE, Francks C. A schizophrenia-associated HLA locus affects thalamus volume and asymmetry. *Brain Behav Immun*. 2015;46:311-318. doi:10.1016/j.bbi.2015.02.021
24. Walters JT, Rujescu D, Franke B, et al. The role of the major histocompatibility complex region in cognition and brain structure: a schizophrenia GWAS follow-up. *Am J Psychiatr*. 2013;170:877-885.
25. Le Clerc S, Taing L, Fond G, et al. A double amino-acid change in the HLA-A peptide-binding groove is associated with response to psychotropic treatment in patients with schizophrenia. *Transl Psychiatry*. 2015;5:e608-e608.
26. Legge SE, Santoro ML, Periyasamy S, Okewole A, Arsalan A, Kowalec K. Genetic architecture of schizophrenia: a review of major advancements. *Psychol Med*. 2021;51:2168-2177.
27. Kodavali CV, Watson AM, Prasad KM, et al. HLA associations in schizophrenia: Are we re-discovering the wheel? *Am J Med Genet A*. 2014;165:19-27. doi:10.1002/ajmg.b.32195
28. Sanchez-Mazas A, Lemaître JF, Currat M. Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. *Philos Trans R Soc Lond B Biol Sci*. 2012;367:830-839.
29. Salamon H, Klitz W, Easteal S, et al. Evolution of HLA class II molecules: Allelic and amino acid site variability across populations. *Genetics*. 1999;152: 393-400.
30. Garamszegi LZ. Global distribution of malaria-resistant MHC-HLA alleles: the number and frequencies of alleles and malaria risk. *Malar J*. 2014;13:349.
31. Singh R, Kaul R, Kaul A, Khan K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J Gastroenterol*. 2007;13:1770-1787. doi:10.3748/wjg.v13.i12.1770
32. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. *Clin Microbiol Rev*. 2009;22:370-385.
33. Hill AV, Yates SN, Allsopp CE, et al. Human leukocyte antigens and natural selection by malaria. *Philos Trans R Soc Lond B Biol Sci*. 1994;346:379-385. doi:10.1098/rstb.1994.0155.
34. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. *J Autoimmun*. 2015;64:13-25. doi:https://doi.org/10.1016/j.jaut.2015.06.010
35. Vlachopoulou E, Lahtela E, Wennerström A, et al. Evaluation of HLA-DRB1 imputation using a Finnish dataset. *Tissue Antigens*. 2014;83:350-355. doi:10.1111/tan.12343
36. James L, Georgopoulos A. The human leukocyte antigen (HLA) DRB1*13:02 allele protects against dementia in continental Western Europe. *J Neurol Neuromed*. 2019;4:1-6. doi:10.29245/2572.942X/2019/5.1253
37. James L, Georgopoulos A. Tri-allelic Human Leukocyte Antigen (HLA) protection against dementia. *J Neurol Neuromed*. 2019;5:12-17. doi:10.29245/2572.942X/2019/1.1261
38. James L, Georgopoulos A. Shared human leukocyte antigen (HLA) coverage in dementia and Parkinson's disease. *J Neurol Neuromed*. 2020;5:45-54. doi:10.29245/2572.942X/2020/3.1275
39. James LM, Georgopoulos AP. Immunogenetic epidemiology of dementia and Parkinson's disease in 14 continental European countries: shared human leukocyte antigen (HLA) profiles. *Journal Immunol Sci*. 2021;5:16-26. doi:10.29245/2578-3009/2021/2.1209
40. James LM, Georgopoulos AP. Immunogenetic epidemiology of multiple sclerosis in 14 continental Western European countries. *J Immunol Sci*. 2021;5:40-46. doi:10.29245/2578.3009/2021/2.1216
41. James LM, Georgopoulos AP. Immunogenetic epidemiology of motor neuron diseases in 14 continental Western European countries. *J Immunol Sci*. 2021;5:22-28. doi:http://dx.doi.org/10.29245/2578-3009/2021/3.1221
42. James LM, Georgopoulos AP. Immunogenetic epidemiology of Type 1 diabetes in 14 Continental Western European countries. *J Immunol Sci*. 2021;5:29-35. doi:10.29245/2578-3009/2021/3.1219
43. Georgopoulos AP, James LM. Immunogenetic association of 127 Human Leukocyte Antigen (HLA) alleles with 30 cancers in continental Western European countries. *Journal Immunol Sci*. 2022;6:6-17. doi:10.29245/2578-3009/2022/2.1227
44. James LM, Georgopoulos AP. Protective and susceptibility wffects of Human Leukocyte Antigen on melanoma prevalence and their implications for predicting checkpoint blockade immunotherapy. *J Immunol Sci*. 2022;6:25-35.
45. James LM, Georgopoulos AP. Immunogenetic profiles and associations of breast, cervical, ovarian, and uterine cancers. *Cancer Inform*. 2023;22: 2211769351221148588. doi:10.1177/11769351221148588
46. Fainboim L, Cañero Velasco MC, Marcos CY, et al. Protracted, but not acute, hepatitis A virus infection is strongly associated with HLA-DRB*1301, a marker for pediatric autoimmune hepatitis. *Hepatology*. 2001;33:1512-1517. doi:10.1053/jhep.2001.24562
47. Brown AS, Derkits EJ. Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. *Am J Psychiatr*. 2010;167:261-280.
48. Khandaker GM, Zimbron J, Dalman C, Lewis G, Jones PB. Childhood infection and adult schizophrenia: a meta-analysis of population-based studies. *Schizophr Res*. 2012;139:161-168.
49. Khandaker GM, Zimbron J, Lewis G, Jones PB. Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol Med*. 2013;43:239-257. doi:10.1017/S0033291712000736
50. Arias I, Sorlozano A, Villegas E, et al. Infectious agents associated with schizophrenia: a meta-analysis. *Schizophr Res*. 2012;136:128-136.
51. Staras SA, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. Seroprevalence of cytomegalovirus infection in the United States, 1988-1994. *Clin Infect Dis*. 2006;43:1143-1151. doi:10.1086/508173.51
52. Wald A, Corey L. Persistence in the population: epidemiology, transmission. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al., eds. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*. Cambridge University Press; 2007. https://www.ncbi.nlm.nih.gov/books/NBK47447
53. McQuillan G, Kruszon-Moran D, Flagg EW, Paulose-Ram R. *Prevalence of herpes simplex virus type 1 and type 2 in persons aged 14–49: United States, 2015–2016*. NCHS Data Brief, no 304. Hyattsville, MD: National Center for Health Statistics; 2018.
54. Yolken RH, Torrey EF. Viruses, schizophrenia, and bipolar disorder. *Clin Microbiol Rev*. 1995;8:131-145.
55. Torrey EF, Leweke MF, Schwarz MJ, et al. Cytomegalovirus and schizophrenia. *CNS Drugs*. 2006;20:879-885.
56. Prasad KM, Shirts BH, Yolken RH, Keshavan MS, Nimgaonkar VL. Brain morphological changes associated with exposure to HSV1 in first-episode schizophrenia. *Mol Psychiatry*. 2007;12:105-11313, 1. doi:10.1038/sj.mp.4001915
57. Dickerson F, Origoni A, Schweinfurth LAB, et al. Clinical and serological predictors of suicide in schizophrenia and major mood disorders. *J Nerv Ment Dis*. 2018;206:173-178.
58. Dickerson FB, Boronow JJ, Stallings C, Origoni AE, Ruslanova I, Yolken RH. Association of serum antibodies to herpes simplex virus 1 with cognitive deficits in individuals with schizophrenia. *Arch Gen Psychiatry*. 2003;60:466-472.
59. Shirts B, Prasad K, Pogue-Geile M, Dickerson F, Yolken R, Nimgaonkar V. Antibodies to cytomegalovirus and herpes simplex virus 1 associated with cognitive function in schizophrenia. *Schizophr Res*. 2008;106:268-274. doi:10.1016/j.schres.2008.07.011
60. Dickerson FB, Boronow JJ, Stallings CR, Origoni AE, Yolken RH. Reduction of symptoms by valacyclovir in cytomegalovirus-seropositive individuals with schizophrenia. *Am J Psychiatr*. 2003;160:2234-2236.
61. Prasad KM, Watson AM, Dickerson FB, Yolken RH, Nimgaonkar VL. Exposure to herpes simplex virus type 1 and cognitive impairments in individuals with schizophrenia. *Schizophr Bull*. 2012;38:1137-1148.
62. Müller N. Inflammation in schizophrenia: pathogenetic aspects and therapeutic considerations. *Schizophr Bull*. 2018;44:973-982.
63. Kim JJ, Shirts BH, Dayal M, et al. Are exposure to cytomegalovirus and genetic variation on chromosome 6p joint risk factors for schizophrenia? *Ann Med*. 2007;39:145-153.
64. Shirts BH, Kim JJ, Reich S, et al. Polymorphisms in MICB are associated with human herpes virus seropositivity and schizophrenia risk. *Schizophr Res*. 2007;94: 342-353.

65. James L, Georgopoulos A. Persistent antigens hypothesis: the human leukocyte antigen (HLA) connection. *JNeurolNeuromed* 2018;3:27-31. doi:10.29245/2572.942X/2018/6.1235
66. James LM, Georgopoulos AP. Human leukocyte antigen as a key factor in preventing dementia and associated apolipoprotein E4 risk. *Front Aging Neurosci*. 2019;11:82.
67. Charonis S, James LM, Georgopoulos AP. *In silico* assessment of binding affinities of three dementia-protective human leukocyte antigen (HLA) alleles to nine human herpes virus antigens. *Curr Res Transl Med*. 2020;68:211-216. doi:10.1016/j.retram.2020.06.002
68. James LM, Charonis SA, Georgopoulos AP. Association of dementia Human Leukocyte Antigen (HLA) profile with human herpes viruses 3 and 7: an *in-silico investigation*. *Journal Immunol Sci*. 2021;5:7-14. doi:10.29245/2578-3009/2021/3.1218
69. Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2016 (GBD 2016) Results*. Seattle: Institute for Health Metrics and Evaluation (IHME), 2020. Accessed July 5, 2021. http://ghdx.healthdata.org/gbd-results-tool.
70. Population Reference Bureau. *2016 world population data sheet with a special focus on human needs and sustainable resources*. Population Reference Bureau, Washington, DC, 2016. Accessed February 5, 2019. https://www.prb.org/2016-world-population-data-sheet/
71. The Uniprot Consortium. 2019. https://www.uniprot.org/
72. Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol*. 2013;31:443-473.
73. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020;48:W449-W454. doi:10.1093/nar/gkaa379
74. Tukey JW. *Exploratory Data Analysis*. Pearson; 1977.
75. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Erlbaum; 1988.
76. Cohen J. A power primer. *Psychol Bull*. 1992;112:155-159.
77. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Chapman & Hall; 1993.
78. Price GR. Selection and covariance. *Nature*. 1970;227:520-521. doi:10.1038/227520a0
79. Frank SA. George Price's contributions to evolutionary genetics. *J Theor Biol*. 1995;175:373-388.
80. Markowitz H. Portfolio selection. *J Finance*. 1952;7:77-91.
81. Childs EJ, Sobel EM, Palmer CG, Sinsheimer JS. Detection of intergenerational genetic effects with application to HLA-B matching as a risk factor for schizophrenia. *Hum Hered*. 2011;72:161-172.
82. Sheehan NJ. The ramifications of HLA-B27. *J R Soc Med*. 2004;97:10-14.
83. Sellgren C, Frisell T, Lichtenstein P, Landen M, Askling J. The association between schizophrenia and rheumatoid arthritis: A nationwide population-based Swedish study on intraindividual and Familial Risks. *Schiz Bull*. 2014;40:1552-1559. doi:10.1093/schbul/sbu054
84. Eaton WW, Byrne M, Ewald H, et al. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatr*. 2006;163:521-528.
85. Benros ME, Mortensen PB, Eaton WW. Autoimmune diseases and infections as risk factors for schizophrenia. *Ann N Y Acad Sci*. 2012;1262:56-66. doi:10.1111/j.1749-6632.2012.06638.x
86. Benros ME, Nielsen PR, Nordentoft M, Eaton WW, Dalton SO, Mortensen PB. Autoimmune diseases and severe infections as risk factors for schizophrenia: a 30-year population-based register study. *Am J Psychiatr*. 2011;168:1303-1310.
87. Benros ME, Pedersen MG, Rasmussen H, Eaton WW, Nordentoft M, Mortensen PB. A nationwide study on the risk of autoimmune diseases in individuals with a personal or a family history of schizophrenia and related psychosis. *Am J Psychiatr*. 2014;171:218-226.
88. Chen SJ, Chao YL, Chen CY, et al. Prevalence of autoimmune diseases in in-patients with schizophrenia: nationwide population-based study. *Br J Psychiatry*. 2012;200:374-380.
89. Ungprasert P, Wijarnpreecha K, Cheungpasitporn W. Patients with psoriasis have a higher risk of schizophrenia: a systematic review and meta-analysis of observational studies. *J Postgrad Med*. 2019;65:141-145.
90. Kirkpatrick B, Miller BJ. Inflammation and schizophrenia. *Schizophr Bull*. 2013;39:1174-1179.
91. Feigenson KA, Kusnecov AW, Silverstein SM. Inflammation and the two-hit hypothesis of schizophrenia. *Neurosci Biobehav Rev*. 2014;38:72-93.
92. Khandaker GM, Cousins L, Deakin J, Lennox BR, Yolken R, Jones PB. Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. *Lancet Psychiatry*. 2015;2:258-270.
93. Al-Diwani AAJ, Pollak TA, Irani SR, Lennox BR. Psychosis: an autoimmune disease? *Immunology*. 2017;152:388-401. doi:10.1111/imm.12795
94. Meehl PE. Schizotaxia, schizotypy, schizophrenia. *Am Psychol*. 1962;17:827-838.
95. Kwapil TR, Barrantes-Vidal N. Schizotypy: looking back and moving forward. *Schizophr Bull*. 2015;41 Suppl 2(Suppl 2):S366-S373. doi:10.1093/schbul/sbu186
96. Barrantes-Vidal N, Grant P, Kwapil TR. The role of schizotypy in the study of the etiology of schizophrenia spectrum disorders. *Schizophr Bull*. 2015;41(Suppl 2):S408-S416. doi:10.1093/schbul/sbu191
97. Fonseca-Pedrero E, Debbané M, Rodríguez-Testal JF, Cohen AS, Docherty AR, Ortuño-Sierra J. Schizotypy: the way ahead. *Psicothema*. 2021;33:16-27.
98. Mohagheghi M, Eftekharian MM, Taheri M, Alikhani MY. Determining the IgM and IgG antibodies titer against HSV1, HSV2 and CMV in the serum of schizophrenia patients. *Hum Antibodies*. 2019;26:87-93. doi:10.3233/HAB-170325
99. Delisi LE, Smith SB, Hamovit JR, et al. Herpes simplex virus, cytomegalovirus and Epstein-barr virus antibody titres in sera from schizophrenic patients. *Psychol Med*. 1986;16:757-763. doi:10.1017/S0033291700011764
100. Schretlen DJ, Vannorsdall TD, Winicki JM, et al. Neuroanatomic and cognitive abnormalities related to herpes simplex virus type 1 in schizophrenia. *Schizophr Res*. 2010;118:224-231.
101. Tucker JD, Bertke AS. Assessment of cognitive impairment in HSV-1 positive schizophrenia and bipolar patients: systematic review and meta-analysis. *Schizophr Res*. 2019;209:40-47. doi:10.1016/j.schres.2019.01.001
102. Dickerson F, Schroeder JR, Nimgaonkar V, Gold J, Yolken R. The association between exposure to herpes simplex virus type 1 (HSV-1) and cognitive functioning in schizophrenia: A meta-analysis. *Psychiatry Res*. 2020;291:113157. doi:10.1016/j.psychres.2020.113157
103. Charonis S, James L, Georgopoulos A. In silico analysis of the binding affinities of antigenic epitopes of vaccines administered to Gulf war veterans to specific HLA Class II alleles protective for Gulf war Illness. *J Neurol Neuromed*. 2019;4:23-30. doi:10.29245/2572.942X/2019/5.1254
104. Charonis SA, James LM, Georgopoulos AP. SARS-CoV-2 in silico binding affinity to human leukocyte antigen (HLA) Class II molecules predicts vaccine effectiveness across variants of concern (VOC). *Sci Rep*. 2022;12:18074-18110. doi:10.1038/s41598-022-11956-5
105. James L, Tsilibary EP, Charonis S, Georgopoulos A. Commentary: COVID-19 and the path to Immunity. *Journal Immunol Sci*. 2020;4:31-33. doi:10.29245/2578-3009/2020/4.1201
106. James LM, Georgopoulos AP. At the root of 3 "long" diseases: persistent antigens inflicting chronic damage on the brain and other organs in Gulf war illness, Long-COVID-19, and chronic fatigue syndrome. *Neurosci Insights*. 2022;17:26331055221114817.
107. Griffin DE. Why does viral RNA sometimes persist after recovery from acute infections? *PLoS Biol*. 2022;20:e3001687.
108. Yousuf W, Ibrahim H, Harfouche M, Abu-Hijleh F, Abu-Raddad L. Herpes simplex virus type 1 in Europe: systematic review, meta-analyses and meta-regressions. *BMJ Glob Health*. 2020;5:e002388.
109. Brainstorm Consortium. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360:eaap8757.
110. Zhang R, Sjölander A, Ploner A, Lu D, Bulik CM, Bergen SE. Novel disease associations with schizophrenia genetic risk revealed in ~400,000 UK Biobank participants. *Mol Psychiatry*. 2022;27:1448-1454.
111. Zheutlin AB, Dennis J, Karlsson Linnér R, et al. Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. *Am J Psychiatr*. 2019;176:846-855.
112. Oliveira J, Oliveira-Maia AJ, Tamouza R, Brown AS, Leboyer M. Infectious and immunogenetic factors in bipolar disorder. *Acta Psychiatr Scand*. 2017;136:409-423. doi:10.1111/acps.12791
113. Tamouza R, Oliveira J, Etain B, et al. HLA genetics in bipolar disorder. *Acta Psychiatr Scand*. 2018;138:464-471.
114. Rosenblat JD. Targeting the immune system in the treatment of bipolar disorder. *Psychopharmacology*. 2019;236:2909-2921.
115. James C, Harfouche M, Welton NJ, et al. Herpes simplex virus: global infection prevalence and incidence estimates, 2016. *Bull World Health Organ*. 2020;98:315-329.

Appendix

Amino acid sequences of the 9 HHV strains (Table 4). Strain labels are from Uniprot (https://www.uniprot.org/uniprotkb/)

HHV1: Q69091 · GD_HHV11	Envelope glycoprotein D	394 AA

MGGAAARLGAVILFVVIVGLHGVRSKYALVDASLKMADPNRFRGKDLPVLDQLTDPPGVRRVYHIQAGLPDPFQPPSLPITVYYAVLERACRSVLLNAPSEAPQIVRGASEDVRKQPYNLTIA
WFRMGGNCAIPITVMEYTECSYNKSLGACPIRTQPRWNYYDSFSAVSEDNLGFLMHAPAFETAGTYLRLVKINDWTEITQFILEHRAKGSCKYALPLRIPPSACLSPQAYQQGVTVDSIGMLP
RFIPENQRTVAVYSLKIAGWHGPKAPYTSTLLPPELSETPNATQPELAPEDPEDSALLEDPVGTVAPQIPPNWHIPSIQDAATPYHPPATPNNMGLIAGAVGGSLLAALVICGIVYWMRRHTQ
KAPKRIRLPHIREDDQPSSHQPLFY

HHV2: P03172 · GD_HHV23	Envelope glycoprotein D	393 AA

MGRLTSGVGTAALLVVAVGLRVVCAKYALADPSLKMADPNRFRGKNLPVLDRLTDPPGVKRVYHIQPSLEDPFQPPSIPITVYYAVLERACRSVLLHAPSEAPQIVRGASDEARKHTYNLTIA
WYRMGDNCAIPITVMEYTECPYNKSLGVCPIRTQPRWSYYDSFSAVSEDNLGFLMHAPAFETAGTYLRLVKINDWTEITQFILEHRARASCKYALPLRIPPAACLTSKAYQQGVTVDSIGMLP
RFIPENQRTVALYSLKIAGWHGPKPPYTSTLLPPELSDTTNATQPELVPEDPEDSALLEDPAGTVSSQIPPNWHIPSIQDVAPHHAPAAPSNPGLIIGALAGSTLAVLVIGGIAFWVRRRAQM
APKRLRLPHIRDDDAPPSHQPLFY

HHV3: Q9J3M8 · GE_VZVO	Envelope glycoprotein E	623 AA

MGTVNKPVVGVLMGFGIITGTLRITNPVRASVLRYDDFHIDEDKLDTNSVYEPYYHSDHAESSWVNRGESSRKAYDHNSPYIWPRNDYDGFLENAHEHHGVYNQGRGIDSGERLMQPTQMSAQ
EDLGDDTGIHVIPTLNGDDRHKIVNVDQRQYGDVFKGDLNPKPQGQRLIEVSVEENHPFTLRAPIQRIYGVRYTETWSFLPSLTCTGDAAPAIQHICLKHTTCFQDVVVDVDCAENTKEDQLA
EISYRFQGKKEADQPWIVVNTSTLFDELELDPPEIEPGVLKVLRTEKQYLGVYIWNMRGSDGTSTYATFLVTWKGDEKTRNPTPAVTPQPRGAEFHMWNYHSHVFSVGDTFSLAMHLQYKIHE
APFDLLLEWLYVPIDPTCQPMRLYSTCLYHPNAPQCLSHMNSGCTFTSPHLAQRVASTVYQNCEHADNYTAYCLGISHMEPSFGLILHDGGTTLKFVDTPESLSGLYVFVVYFNGHVEAVAYT
VVSTVDHFVNAIEERGFPPTAGQPPATTKPKEITPVNPGTSPLLRYAAWTGGLAAVVLLCLVIFLICTAKRMRVKAYRVDKSPYNQSMYYAGLPVDDFEDSESTDTEEEFGNAIGGSHGGSSY
TVYIDKTR

HHV4: P03188 · GB_EBVB9	Envelope glycoprotein B	857 AA

MTRRRVLSVVVLLAALACRLGAQTPEQPAPPATTVQPTATRQQTSFPFRVCELSSHGDLFRFSSDIQCPSFGTRENHTEGLLMVFKDNIIPYSFKVRSYTKIVTNILIYNGWYADSVTNRHEE
KFSVDSYETDQMDTIYQCYNAVKMTKDGLTRVYVDRDGVNITVNLKPTGGLANGVRRYASQTELYDAPGWLIWTYRTRTTVNCLITDMMAKSNSPFDFFVTTTGQTVEMSPFYDGKNKETFHE
RADSFHVRTNYKIVDYDNRGTNPQGERRAFLDKGTYTLSWKLENRTAYCPLQHWQTFDSTIATETGKSIHFVTDEGTSSFVTNTTVGIELPDAFKCIEEQVNKTMHEKYEAVQDRYTKGQEAI
TYFITSGGLLLAWLPLTPRSLATVKNLTELTTPTSSPPSSPSPPAPSAARGSTPAAVLRRRRRDAGNATTPVPPTAPGKSLGTLNNPATVQIQFAYDSLRRQINRMLGDLARAWCLEQKRQNM
VLRELTKINPTTVMSSIYGKAVAAKRLGDVISVSQCVPVNQATVTLRKSMRVPGSETMCYSRPLVSFSFINDTKTYEGQLGTDNEIFLTKKMTEVCQATSQYYFQSGNEIHVYNDYHHFKTIE
LDGIATLQTFISLNTSLIENIDFASLELYSRDEQRASNVFDLEGIFREYNFQAQNIAGLRKDLDNAVSNGRNQFVDGLGELMDSLGSVGQSITNLVSTVGGLFSSLVSGFISFFKNPFGGMLI
LVLVAGVVILVISLTRRTRQMSQQPVQMLYPGIDELAQQHASGEGPGINPISKTELQAIMLALHEQNQEQKRAAQRAAGPSVASRALQAARDRFPGLRRRRYHDPETAAALLGEAETEF

HHV5: P06473 · GB_HCMVA	Envelope glycoprotein B	gB	906 AA

MESRIWCLVVCVNLCIVCLGAAVSSSSTSHATSSTHNGSHTSRTTSAQTRSVYSQHVTSSEAVSHRANETIYNTTLKYGDVVGVNTTKYPYRVCSMAQGTDLIRFERNIICTSMKPINEDLDE
GIMVVYKRNIVAHTFKVRVYQKVLTFRRSYAYIYTTYLLGSNTEYVAPPMWEIHHINKFAQCYSSYSRVIGGTVFVAYHRDSYENKTMQLIPDDYSNTHSTRYVTVKDQWHSRGSTWLYRETC
NLNCMLTITTARSKYPYHFFATSTGDVVYISPFYNGTNRNASYFGENADKFFIFPNYTIVSDFGRPNAAPETHRLVAFLERADSVISWDIQDEKNVTCQLTFWEASERTIRSEAEDSYHFSSA
KMTATFLSKKQEVNMSDSALDCVRDEAINKLQQIFNTSYNQTYEKYGNVSVFETSGGLVVFWQGIKQKSLVELERLANRSSLNITHRTRRSTSDNNTTHLSSMESVHNLVYAQLQFTYDTLRG
YINRALAQIAEAWCVDQRRTLEVFKELSKINPSAILSAIYNKPIAARFMGDVLGLASCVTINQTSVKVLRDMNVKESPGRCYSRPVVIFNFANSSYVQYGQLGEDNEILLGNHRTEECQLPSL
KIFIAGNSAYEYVDYLFKRMIDLSSISTVDSMIALDIDPLENTDFRVLELYSQKELRSSNVFDLEEIMREFNSYKQRVKYVEDKVVDPLPPYLKGLDDLMSGLGAAGKAVGVAIGAVGGAVAS
VVEGVATFLKNPFGAFTIILVAIAVVIITYLIYTRQRRLCTQPLQNLFPYLVSADGTTVTSGSTKDTSLQAPPSYEESVYNSGRKGPGPPSSDASTAAPPYTNEQAYQMLLALARLDAEQRAQ
QNGTDSLDGQTGTQDKGQKPNLLDRLRHRKNGYRHLKDSDEEENV

HHV6A: P0DOE0 · GQ2_HHV6U	Envelope glycoprotein Q2	214 AA

MHFLVVYILIHFHAYRGMAALPLFSTLPKITSCCDSYVVINSSTSVSSLISTCLDGEILFQNEGQKFCRPLTDNRTIVYTMQDQVQKPLSVTWMDFNLVISDYGRDVINNLTKSAMLARKNGP
RYLQMENGPRYLQMETRISDLFRHECYQDNYYVLDKKLQMFYPTTHSNELLFYPSEATLPSPWQEPPFSSPWPEPTFPSRWYWLLLNYTNY

HHV6B: Q9QJ11 · GQ1_HHV6Z	Envelope glycoprotein Q1	516 AA

MRPPRRSAPILVCAISMATALSNATVHRDAGTVESTPPPDDEDNYTAKYYDDSIYFNIYDGTNPTPRRRTLPEIISKFSTSEMSRLGGLKAFVPVDYTPTTTLEDIEDLLNYAICDDNSCGCL
IETEARXMFGDIIICVPLSAESRGVRNLKSRIMPMGLSQILSSGLGLHFSLLYGAFGSNYNSLAYMERLKPLTAMTAIAFCPMTSKLELRQNYRLEKARXNLIVNIELLKIQNHGGQTIKTLT
SFAIVRKDSDGQDWETCTRFASVSIEDILRSKPAANGTCCPPRDVHHDRPTLQSSNSWTRTEYFEPWQDVVDAYVPINDNHCPNDSYVVFQTLQGHEWCSRLNKNDTKNYLSSVLAFKNALYE
TEELMETIGMRLASQILSLVGQRGTSIRNIDPAIVSALWHSLPEKLTTTNIKYDIASPTHMSPALXTIFIQTGTSKQRFRNAGLLMVNNIFTVQARYSKQNMFEKKIYGYEHLGQALCEGGHV
FYNPRDVYFQNIKMAATEPTVVRT

HHV7: P52353 · GH_HHV7J	Envelope glycoprotein H	690 AA

MYFYINSLLLIVSINGWKHWNILNSSICVNEKTNQTIIQPGLITFNFHDYNETRVYQIPKCLFGYTFVSNLFDSVNFDESFDQYKHRITRFFNPSTEKAVKIYAQKFQTNIKPVSHTKTITVS
FLPLFYEKDVYFANVSEIRKLYYNQYICTLSNGLTDYLFPITERCVMRHYNYLNTVFMLALTPSFFIISVETGMDDVVFIFGNVSRIFFKAPFRKSSFIYRQTVSDDLLLITKKTTIERFYPF
LKIDFLDDIWKQNYDISFLIAKFNKLATVYIMEGFCGKPVNKDTFHLMFLFGLTHFLYSTRGDGLLPLLEILNTHQSIITMGRFLEKCFKMTKSHLLYPEMEKLQNFQLVDYSYITSDLTIPI
SAKLAFLSLADGRIVTVPQNKWKEIENNIETLYEKHKLFTNLTQPERANLFLLSEIGNSLVFQEKIKRKIHVLLASLCNPLEMYFWTHMLDNVMDIETMFSPCATATRKDLTQRVVNNILSYK
NLDAYTNKVMNTLSVYRKKRLDMFKSISCVSNEQAAFLTLPNITYTISSKYILAGTSFSVTSTVISTTIIITVVPLNSTCTPTNYKYSVKNIKPIYNISSHDCVFCESLVVEYDDIDGIIQFV
YIMDDKQLLKLIDPDTNFIDVNPRTHYLLFLRNGSVFEITALDLKSSQVSIMLVLLYLIIIIIVLFGIYHVFRLF

HHV8: F5HAK9 · GH_HHV8P	Envelope glycoprotein H	730 AA

MQGLAFLAALACWRCISLTCGATGALPTTATTITRSATQLINGRTNLSIELEFNGTSFFLNWQNLLNVITEPALTELWTSAEVAEDLRVTLKKRQSLFFPNKTVVISGDGHRYTCEVPTSSQT
YNITKGFNYSALPGHLGGFGINARLVLGDIFASKWSLFARDTPEYRVFYPMNVMAVKFSISIGNNESGVALYGVVSEDFVVVTLHNRSKEANETASHLLFGLPDSLPSLKGHATYDELTFARN
AKYALVAILPKDSYQTLLTENYTRIFLNMTESTPLEFTRTIQTRIVSIEARRACAAQEAAPDIFLVLFQMLVAHFLVARGIAEHRFVEVDCVCRQYAELYFLRRISRLCMPTFTTVGYNHTTL
GAVAATQIARVSATKLASLPRSSQETVLAMVQLGARDGAVPSSILEGIAMVVEHMYTAYTYVYTLGDTERKLMLDIHTVLTDSCPPKDSGVSEKLLRTYLMFTSMCTNIELGEMIARFSKPDS
LNIYRAFSPCFLGLRYDLHPAKLRAEAPQSSALTRTAVARGTSGFAELLHALHLDSLNLIPAINCSKITADKIIATVPLPHVTYIISSEALSNAVVYEVSEIFLKSAMFISAIKPDCSGFNFS
QIDRHIPIVYNISTPRRGCPLCDSVIMSYDESDGLQSLMYVTNERVQTNLFLDKSPFFDNNNLHIHYLWLRDNGTVVEIRGMYRRRAASALFLILSFIGFSGVIYFLYRLFSILY